UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO

DANIELLE TEIXEIRA DE OLIVEIRA

METADADOS PARA REPRESENTAÇÃO DE DADOS ABERTOS: uma proposta para o Portal Brasileiro de Dados Abertos

Belo Horizonte
2025

Danielle Teixeira de Oliveira

METADADOS PARA REPRESENTAÇÃO DE DADOS ABERTOS: uma proposta para o Portal Brasileiro de Dados Abertos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento (PPGGOC), Escola de Ciência da Informação da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Mestre, área de concentração Ciência da Informação.

Linha de Pesquisa: Gestão e Tecnologia da Informação e Comunicação (GETIC)

Orientadora: Profa. Dra. Patrícia Nascimento Silva

Belo Horizonte
2025

O48m
T

Oliveira, Danielle Teixeira de, 1994-
Metadados para representação de dados abertos [recurso eletrônico]: uma proposta para o Portal Brasileiro de Dados Abertos / Danielle Teixeira de Oliveira. - Belo Horizonte, 2025.
1 recurso online (221 f.: il., col.) pdf.

Dissertação -- (Mestrado) - Universidade Federal de Minas Gerais, Escola de Ciência da Informação.
Orientadora: Patrícia Nascimento Silva
Bibliografia: f. 179-197
Apêndices: f. 198-212
Anexos: f. 213-221

1. Ciência da Informação -- Teses. 2. Metadados -- Teses. 3. Catalogação descritiva. 4. Dados. 5. Organização da Informação.
I. Título. II. Silva, Patrícia Nascimento. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDD- 025.32

Bibliotecária: Danielle Teixeira de Oliveira - CRB6 3516/O

Como citar: OLIVEIRA, Danielle Teixeira de. **Metadados para representação de dados abertos**: proposta para o Portal Brasileiro de Dados Abertos. 2025. 221 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2025.

UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPGGOC

FOLHA DE APROVAÇÃO

METADADOS PARA REPRESENTAÇÃO DE DADOS ABERTOS: uma proposta para o Portal Brasileiro de Dados Abertos

DANIELLE TEIXEIRA DE OLIVEIRA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Gestão e Tecnologia da Informação e Comunicação.

Aprovada em 17 de dezembro de 2024, por videoconferência, pela banca constituída pelos membros:

Prof(a). Patricia Nascimento Silva (Orientadora)
ECI/UFMG

Prof(a). Célia da Consolação Dias
ECI/UFMG

Prof(a). Cíntia de Azevedo Lourenço
ECI/UFMG

Prof(a). Janicy Aparecida Pereira Rocha
Unirio

Belo Horizonte, 17 de dezembro de 2024.

Documento assinado eletronicamente por **Patricia Nascimento Silva**, **Professora do Magistério Superior**, em 10/01/2025, às 15:48, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Cintia de Azevedo Lourenco**, **Coordenador(a) de curso**, em 22/01/2025, às 13:30, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Celia da Consolação Dias**, **Professora do Magistério Superior**, em 23/01/2025, às 09:30, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Janicy Aparecida Pereira Rocha**, **Usuário Externo**, em 23/01/2025, às 15:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3775378** e o código CRC **C40379C1**.

Referência: Processo nº 23072.270190/2024-68 SEI nº 3775378

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPGGOC

ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA

DANIELLE TEIXEIRA DE OLIVEIRA

Realizou-se, no dia 17 de dezembro de 2024, às 14:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *METADADOS PARA REPRESENTAÇÃO DE DADOS ABERTOS: uma proposta para o Portal Brasileiro de Dados Abertos*, apresentada por DANIELLE TEIXEIRA DE OLIVEIRA, número de registro 2023651918, graduada no curso de BIBLIOTECONOMIA, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Patricia Nascimento Silva - ECI/UFMG (Orientadora), Prof(a). Célia da Consolação Dias - ECI/UFMG, Prof(a). Cíntia de Azevedo Lourenço - ECI/UFMG, Prof(a). Janicy Aparecida Pereira Rocha - Unirio.

A Comissão considerou a dissertação:

(X) Aprovada

( ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
Belo Horizonte, 17 de dezembro de 2024.

Assinatura dos membros da banca examinadora:

Documento assinado eletronicamente por **Patricia Nascimento Silva**, **Professora do Magistério Superior**, em 10/01/2025, às 15:45, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Cintia de Azevedo Lourenco**, **Coordenador(a) de curso**, em 22/01/2025, às 13:30, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Celia da Consolação Dias**, **Professora do Magistério Superior**, em 23/01/2025, às 09:30, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Janicy Aparecida Pereira Rocha**, **Usuário Externo**, em 23/01/2025, às 15:50, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3775347** e o código CRC **892E2272**.

Referência: Processo nº 23072.270190/2024-68 SEI nº 3775347

Aos meus filhos, Gabriel (in memoriam), por quem meu céu é mais azul, e Manuela, por quem minha vida é mais colorida.
Aos meus pais, Gleice e Henrique, que seguraram minhas mãos nos primeiros passos e confiaram nos caminhos que escolhi seguir.
À minha irmã Dayanne, para que saiba que sempre andaremos juntas, mesmo que por trilhas diferentes.

AGRADECIMENTOS

À fé, por ter me erguido todas as vezes que esmoreci.

À minha filha, Manuela, por ser minha âncora, preservando minha humanidade e me lembrando diariamente que o mais importante eu já tenho.

Ao meu filho Gabriel, por ser minhas asas e me manter próxima ao céu.

À minha família, cujo amor foi fonte de acolhimento e conforto.

Ao Matheus, pai dos meus filhos, por estar com minha filha em todos os momentos que eu não pude durante essa trajetória.

À minha orientadora, Profa. Dra. Patrícia, que me guiou nos trilhos da pesquisa científica como uma excelente tutora.

Aos meus amigos, por terem estado comigo antes e durante o percurso desse sonho, me apoiando e me dando suporte, me ajudando a aguentar, quando eu não aguentava mais.

À UFMG, que me acolhe desde 2014 como estudante e como profissional, sempre me dando oportunidades de evoluir.

Às professoras Dra. Raquel e Dra. Silvana que me auxiliaram a iniciar na trajetória científica ainda no ensino médio.

Ao PPGGOC e às pessoas que a ele integram, por oportunizarem a minha transformação em uma pesquisadora.

Às professoras Dra. Célia, Dra. Cíntia e Dra. Janicy, por gentilmente aceitarem participar da minha banca e agregarem valor à minha pesquisa.

Às pessoas pretas do programa de extensão da Faculdade de Educação da UFMG, Afirmação na Pós, que não só me mostraram que é possível, mas me ensinaram como tornar possível.

Aos meus colegas de trabalho, que generosamente se organizaram e contribuíram para que eu tivesse um valioso aliado nessa jornada: o tempo.

Aos integrantes da Faculdade de Educação da UFMG, por valorizarem a educação pública, gratuita e de qualidade e demonstrarem isso através de ações concretas.

A todos que contribuíram, direta ou indiretamente, para que hoje eu estivesse um passo à frente do que ontem.

Não, não pares. É graça Divina começar bem. Graça maior é persistir na caminhada certa, manter o ritmo... Mas a graça das graças é não desistir. Prosseguir firme. Podendo ou não podendo. Caindo embora aos pedaços… Chegar até o fim - Dom Helder Câmara.

RESUMO

O século XXI tem sido marcado pelo desenvolvimento das tecnologias de comunicação e informação e pela velocidade de produção e consumo de dados. Entre os diversos dados disponíveis na Web, existem os dados abertos, que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa, subdivididos em vários tipos, incluindo os dados governamentais abertos (DGA), que são aqueles originados das funções administrativa e governamental e considerados pilares da democracia e da transparência. O Brasil demonstrou-se pioneiro em relação à disponibilização de dados abertos do governo que podem ser reutilizados e redistribuídos pela sociedade, no entanto, ainda existem lacunas nas normativas, especialmente relacionadas à representação, à organização e ao acesso a esses dados, cenário em que a Ciência da Informação pode apresentar instrumentos e metodologias para apoiar a governança de dados. Tal contexto motivou esta pesquisa, que buscou responder à questão: quais metadados podem ser utilizados para representar os dados abertos do governo federal brasileiro? O objetivo geral foi propor um modelo de metadados para a representação e a organização da informação no Portal Brasileiro de Dados Abertos. Para tanto, realizou-se uma pesquisa aplicada, exploratória e descritiva para atender aos objetivos específicos: (i) identificar, na literatura acadêmica brasileira, modelos, métricas, *frameworks* e instrumentos similares que abordem a publicação, a avaliação, o ciclo de vida ou o tratamento de dados abertos ou DGA, a fim de identificar abordagens, técnicas e métodos utilizados; (ii) mapear documentos e legislações (leis, decretos, normativas, portarias, resoluções e instrumentos) sobre os dados abertos no governo federal brasileiro; (iii) identificar e analisar as tipologias e os formatos de dados abertos publicados, padrões e metadados utilizados no Portal Brasileiro de Dados Abertos e (iv) propor uma estrutura de representação descritiva para os dados abertos do governo federal brasileiro no Portal Brasileiro de Dados Abertos com a utilização de metadados. Para alcançar o objetivo (i), realizou-se uma revisão integrativa da literatura nacional na Biblioteca Digital Brasileira de Teses e Dissertações que resultou na observação de uma preocupação unívoca com os desafios decorrentes da falta de padronização e heterogeneidade dos dados, além da identificação de boas práticas para o tratamento dos dados abertos. No objetivo (ii), foram mapeadas as principais documentações e instrumentos relacionados à pesquisa, como o Padrão de Metadados do Governo Eletrônico e o Vocabulário Controlado de Governo Eletrônico. No objetivo (iii), foi realizada uma pesquisa documental e uma observação direta no Portal Brasileiro de Dados Abertos e, como resultados, foram identificados os elementos de organização e representação da informação no Portal. Por fim, para o objetivo (iv), os resultados anteriores foram sistematizados para a apresentação de recomendações da estrutura de hierarquização, com divisão em categorias, para o Portal Brasileiro de Dados Abertos, estrutura de metadados e correspondência com o padrão de metadados *Data Catalog Vocabulary* 3. Concluiu-se que as recomendações dessa dissertação podem apoiar mudanças significativas no contexto brasileiro, promovendo a interoperabilidade e a acessibilidade para a efetiva reutilização dos dados pela sociedade, destacando o papel essencial do profissional da informação no desenvolvimento das políticas de dados abertos.

Palavras-chave: dados abertos; Portal Brasileiro de Dados Abertos; representação descritiva; metadados; interoperabilidade.

ABSTRACT

The 21st century has been marked by the development of communication and information technologies and the immeasurable speed of data production and consumption. Among the diverse data available on the web, there is open data that can be freely used, reused and redistributed by anyone, subdivided into various types, including open government data (OGD), which are those originated from administrative and governmental functions and considered pillars of democracy and transparency. Brazil has shown itself to be a pioneer when it comes to making open government data available that can be reused and redistributed by society. However, there are still gaps in the regulations, especially related to the representation, organization and access to this data, a scenario in which Information Science can present tools and methodologies to support data governance. This context motivated this research, which sought to answer the question: what metadata can be used to represent the open data of the Brazilian federal government? The general objective was to propose a metadata model for the representation and organization of information on the Brazilian Open Data Portal. To this end, an applied, exploratory and descriptive study was carried out to meet the specific objectives: (i) to identify, in the Brazilian academic literature, models, metrics, frameworks and similar instruments that address the publication, evaluation, life cycle or treatment of open data or DGA, in order to identify approaches, techniques and methods used; (ii) to map documents and legislation on open data in the Brazilian federal government ((laws, decrees, regulations, ordinances, resolutions and instruments); (iii) to identify and analyze the types and formats of open data published, standards and metadata used on the Brazilian Open Data Portal; and (iv) to propose a descriptive representation structure for the open data of the Brazilian federal government using metadata. To achieve objective (i), an integrative review of the national literature was carried out in the Brazilian Digital Library of Theses and Dissertations, which resulted in the observation of an unequivocal concern with the challenges arising from the lack of standardization and heterogeneity of the data, as well as the identification of good practices for the treatment of open data. In objective (ii), the main documentation and instruments related to the research were mapped, such as the Electronic Government Metadata Standard and the Electronic Government Controlled Vocabulary. In objective (iii), documentary research and direct observation were carried out on the Brazilian Open Data Portal and, as a result, the elements of organization and representation of information on the Portal were identified. Finally, for objective (iv), the previous results were systematized to present recommendations for the hierarchical structure, divided into categories, for the Brazilian Open Data Portal, metadata structure and correspondence with the Data Catalog Vocabulary metadata standard. It was concluded that the recommendations of this dissertation can support significant changes in the Brazilian context, promoting interoperability and accessibility for the effective reuse of data by society, highlighting the essential role of information professionals in the development of open data policies.

Keywords: open data; Brazilian Open Data Portal; descriptive representation; metadata; interoperability.

LISTA DE ILUSTRAÇÕES

Figura 1 - Relações entre as sete classes principais do DCAT-3 60

Figura 2 - Relação entre governança de dados e gestão de dados 62

Figura 3 - Arquitetura para publicação de dados na web 63

Figura 4 - Metadados no Portal Brasileiro de Dados Abertos 81

Figura 5 - Metadados de um Conjunto de Dados 82

Figura 6 - Metadados de um Conjunto de Dados 83

Figura 7 - Etapas da seleção de documentos 87

Gráfico 1 - Distribuição de publicações por ano 94

Figura 8 - Painel de monitoramento de dados abertos 114

Figura 9 - Página inicial do Portal Brasileiro de Dados Abertos 116

Figura 10 - Solicitação de autenticação no Portal 117

Figura 11 - Opções de busca no Portal 118

Figura 12 - Menu lateral de navegação no Portal 118

Figura 13 - Filtros de busca no Portal 119

Figura 14 - Filtro de busca por formato do recurso no Portal 123

Gráfico 2 - Metadados em utilização no Portal 125

Figura 15 - Identificação de metadados em um conjunto de dados no Portal 126

Figura 16 - Relacionamento entre as classes propostas 136

Figura 17 - Metadados da categoria Organização 141

Figura 18 - Metadados da classe Conjunto de Dados 149

Figura 19 - Metadados da categoria Recursos 155

Figura 20 - Metadados da categoria Reúso 160

Figura 21 - Diagrama com classes e propriedades do DCAT-3 164

LISTA DE QUADROS

Quadro 1 - Caracterização da catalogação 36
Quadro 2 - Definições do termo ‘metadados’ 39
Quadro 3 - Definição dos tipos de metadados 40
Quadro 4 - Elementos do DC 45
Quadro 5 - Dados dos elementos do e-PMG 47
Quadro 6 - Namespaces do DCAT-3 51
Quadro 7 - Namespaces não normativos do DCAT-3 52
Quadro 8 - Classes do padrão DCAT-3 55
Quadro 9 - Boas Práticas para Dados na Web 63
Quadro 10 - Delineamento Metodológico da Pesquisa 72
Quadro 11 - Protocolo de Revisão Integrativa de Literatura na BDTD 75
Quadro 12 - Documentos da BDTD selecionados para análise 86
Quadro 13 - Iniciativas de dados abertos no governo federal brasileiro 104
Quadro 14 - Quantidades de Organizações, Conjuntos de dados e Reúsos no Portal 119
Quadro 15 - Metadados da Cartilha Técnica para Publicação de Dados Abertos no Brasil 121
Quadro 16 - Metadados do Portal Brasileiro de Dados Abertos 124
Quadro 17 - Resumo da categoria Organização 139
Quadro 18 - Resumo da categoria Conjuntos de Dados 147
Quadro 19 - Resumo da categoria Recursos 153
Quadro 20 - Resumo da categoria Reúso 158
Quadro 21 - Boas práticas para dados na web relacionados a metadados da proposta 166
Quadro 22 - Proposta de metadados para o Portal Brasileiro de Dados Abertos 196
Quadro 23 - Elementos do e-PMG 1.1 205

LISTA DE ABREVIATURAS E SIGLAS

AACR2 - Código de Catalogação Anglo-americano

ADMS - *Asset Description Metadata Schema*

AGLS - *Australian Government Locator Service*

API - *Application Programming Interface*

BCI - Biblioteconomia e Ciência da Informação

BDTD - Biblioteca Digital Brasileira de Teses e Dissertações

CI - Ciência da Informação

CGU - Controladoria-Geral da União

CKAN - *Comprehensive Kerbal Archive Network*

DCAT-3 - *Data Catalog Vocabulary*

DCMI - *Dublin Core Metadata Initiative*

DERI - *Digital Enterprise Research Institute*

DGA - Dados Governamentais Abertos

eGOV - Governo Eletrônico

e-GMS - *Government Metadata Standard*

e-MAG - Modelo de Acessibilidade de Governo Eletrônico

EMBRATUR - Instituto Brasileiro do Turismo

e-PMG - Padrão de Metadados do Governo Eletrônico

e-PING - Padrões de Interoperabilidade de Governo Eletrônico

EUA - Estados Unidos da América

FGDC - *Federal Geographic Data Committee Metadata*

FOAF - *Friend Of Friends*

GILS - *Government Locator Service*

GLD - *Government Linked Data*

INDA - Infraestrutura Nacional de Dados Abertos

INDE - Infraestrutura Nacional de Dados Espaciais

ISBD - *International Standard Bibliographic Description*

LAI - Lei de Acesso à Informação

LCG - Lista de Categorias de Governo

LGPD - Lei Geral de Proteção de Dados

MARC - *Machine Readable Cataloging*

MCT - Ministério de Ciência e Tecnologia

MIP - Metainformação para Interoperabilidade

MIME - *Multipurpose Internet Mail Extension*

OC - Organização do Conhecimento

OCLC - *Online Computer Library Center*

OGP - *Open Government Partnership*

OI - Organização da Informação

OKFN - *Open Knowledge Foundation*

OSCV - *Open Smart City View*

PDA - Plano de Dados Abertos

PDCA - Plan, Do, Check, Act

RDA - *Resource Description and Access*

RDF - *Resource Description Framework*

SGD - Secretaria de Governo Digital

SISP - Administração dos Recursos de Tecnologia da Informação

SOCs - Sistemas de Organização do Conhecimento

SOCINFO - Programa Sociedade da Informação

SIORG - Sistema de Organização e Inovação Institucional do Governo Federal

SRI - Sistema de Recuperação da Informação

TCI - Termo de Classificação de Informação

TICs - Tecnologias de Informação e Comunicação

UFSC - Universidade Federal de Santa Catarina

USP - Universidade de São Paulo

vCard - *vCard Ontology*

VCGE - Vocabulário de Controlado Governo Eletrônico

W3C - *World Wide Web Consortium*

SUMÁRIO

INTRODUÇÃO **18**
1.1 Problema e Justificativas 21
1.2 Objetivos 23
1.2.1 Objetivo geral 23
1.2.2 Objetivos específicos 24
1.3 Estrutura da dissertação 24
2 DADOS ABERTOS NA CIÊNCIA DA INFORMAÇÃO **25**
2.1 Dados, informação e conhecimento 25
2.2 Open Government Data (Dados Governamentais Abertos) 28
3 FUNDAMENTAÇÃO TEÓRICA **32**
3.1 Organização, representação e recuperação da informação 32
3.2 Representação da informação 34
3.3 Metadados e padrões de metadados 38
3.3.1 Dublin Core (DC) 43
3.3.2 Padrão de Metadados do Governo Eletrônico (e-PMG) 45
3.3.3 Vocabulário Controlado de Governo Eletrônico (VCGE 2.1.0) 47
3.3.4 Data Catalog Vocabulary (DCAT-3) 48
3.4 Governança de Dados 59
3.5 Boas Práticas para Dados na Web (Data on the Web Best Practices) 61
4 PROCEDIMENTOS METODOLÓGICOS **68**
4.1 Caracterização da pesquisa 68
4.2 Etapas da pesquisa 73
4.2.1 Etapa 1: revisão integrativa de literatura 73
4.2.2 Etapa 2: mapeamento do contexto brasileiro 77
4.2.3 Etapa 3: mapeamento no Portal Brasileiro de Dados Abertos 78
4.2.3 Etapa 4: proposição de metadados para representação dos dados abertos do governo federal 82
5 PANORAMA DE DADOS ABERTOS NO BRASIL **85**
5.1 Revisão integrativa de literatura 85

5.1.1 Trabalhos correlatos com a proposta da dissertação 100
5.2 Dados abertos do governo federal brasileiro 102
5.2.1 Iniciativas de dados abertos no governo federal brasileiro 102
5.2.2 Portal Brasileiro de Dados Abertos 114
5.2.3 Tipologias de dados abertos, formatos, padrões e metadados utilizados 121
6 PROPOSTA DE MODELO DE METADADOS - METAOPENGOV 132
6.1 Detalhamento da proposta 133
6.1.1 Estrutura e organização 133
6.1.2 Relacionamentos 134
6.1.3 Metadados 136
6.1.3.1 Organização 138
6.1.3.2 Conjunto de Dados 144
6.1.3.3 Recursos 152
6.1.3.4 Reúso 157
6.1.4 Correspondência com o padrão DCAT-3 161
6.2 Análise da Proposta 166
7 CONCLUSÕES 174
7.1 Contribuições da dissertação 176
7.2 Desafios e limitações 177
7.3 Trabalhos futuros 178
7.4 Considerações finais 178
REFERÊNCIAS 180
APÊNDICE A - Proposta de metadados para o Portal Brasileiro de Dados Abertos 199
APÊNDICE B - Sintaxe Turtle da proposta no padrão DCAT 210
ANEXO A - Elementos do e-PMG 1.1 214

INTRODUÇÃO

A segunda década do século XXI vem sendo intensamente marcada pela produção e o consumo de uma quantidade imensurável de dados e informações, caracterizando o fenômeno recentemente conhecido como *big data*, que se define "pelo volume dos dados gerados e disponíveis, pela velocidade com que estes devem ser tratados e apresentados e pela variedade de fontes onde eles se encontram" (Coneglian; Gonçalez; Santarém Segundo, 2017, p. 2).

Esse cenário vem sendo alimentado ao longo dos anos, desde a década de 1960, no contexto da Guerra Fria, quando surgiu a ARPANET, base tecnológica para *internet*, mas ainda restrita e limitada a ambientes militares e de pesquisa. Em 1989, com a criação da *World Wide Web,* por Tim Berners-Lee (W3C, *s.d*.a, *online*), esse ambiente foi revolucionado, tornando-se um dos maiores instrumentos de conexão e interação da contemporaneidade.

A discussão sobre a conceituação de dados e informação vem desde os primórdios da Ciência da Informação (CI), sendo que não existe um consenso absoluto entre os estudiosos da área. Davenport (2002) considera os dados como simples observações sobre o estado do mundo, e a informação como dados dotados de relevância e propósito. No cenário tecnológico atual, em que esses conceitos estão intrinsecamente relacionados, lidar com esse imensurável volume de dados tornou-se um desafio primário para diversos profissionais que lidam com a informação, sobretudo em formato digital.

A CI, desde sua gênese, debruça-se sobre questões relacionadas à organização e à disseminação da informação, e esse novo contexto representa um estímulo para os profissionais dessa área, que buscam, em suas metodologias de estudo, soluções para as novas questões impostas pela explosão informacional dos últimos anos. Nesse processo evolutivo, Saracevic (1996) afirmou existirem três características constituintes da razão de existência e evolução da CI, sendo elas: sua natureza interdisciplinar, sua ligação inexorável com a tecnologia da informação, em que o imperativo tecnológico impôs a transformação da sociedade em era da informação, e o importante papel da CI, a ser desempenhado por sua forte dimensão social e humana, que ultrapassa a tecnologia.

Cerca de três décadas após a criação da Web, a CI encontra-se imersa em um cenário tecnológico ainda mais extremo e absoluto do que aquele descrito por Saracevic (1996). Estima-se que, hoje em dia, seis bilhões de pessoas estejam conectadas à *internet* por meio do 'www', e isso representa 75% da população mundial (Worldometer, 2024, *online*). Esse cenário é

similar ao panorama brasileiro, no qual 84% das casas já possuem acesso à internet (CGI.br/NIC.br; CeticBR, 2023, *online*).

Na perspectiva da CI, Alvarenga (2003) refletiu que, com o advento e o uso intensivo de novas tecnologias, ocorreu um positivo alvoroço relacionado ao campo do conhecimento, em especial em relação à representação, à armazenagem e à recuperação de informações, áreas que são intensamente relacionadas à cognição humana. A CI passa a ser considerada uma ciência interdisciplinar, além de ter características de uma ciência pura, na medida em que investiga seu objeto sem considerar sua aplicação, e também um componente de uma ciência aplicada, visto que desenvolve produtos e serviços (Borko, 1968).

Assim, a CI é a ciência que se dedica a desvendar as propriedades e o comportamento da informação, as forças que governam o fluxo de informações e os meios de processamento de informações para otimizar sua acessibilidade e usabilidade, incluindo investigações sobre os processos de origem, coleção, disseminação, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização de informações (Taylor, 1966; Borko, 1968).

Considerando que os dados estão diretamente relacionados com a informação, eles também são propriedade de estudo da CI e, entre a enorme variedade de dados disponíveis na Web, existem os dados abertos, definidos como aqueles que podem ser livremente usados, reutilizados e redistribuídos por qualquer um, com, no máximo, uma exigência de atribuição de fonte e compartilhamento com as mesmas regras (*Open Knowledge* Brasil, s.d.b). Os próprios dados abertos também são subdivididos em vários tipos, podendo ser dados científicos, financeiros, culturais, ambientais e governamentais, sendo esta última tipologia abordada mais profundamente no presente trabalho.

Considera-se que os dados abertos são essenciais para promover transparência e engajamento social (*Open Knowledge* Brasil, s.d.a). Esse propósito incorpora-se aos dados abertos do governo, dados abertos governamentais ou dados governamentais abertos (DGA), que são aqueles originados das funções administrativa e governamental e que estão disponíveis em formato aberto, não proprietário, não discriminatório e com licenças abertas que garantam a sua reutilização (Pinho, 2021).

Esses dados podem tornar-se acessíveis como DGA a partir de informações orçamentárias, geográficas, sociais, demográficas, legislativas, ambientais, meteorológicas, entre

outras, que são produzidas rotineiramente na execução das atividades públicas dos diversos setores e órgãos do governo e são de valor e interesse público (Pinho, 2021).

Segundo Bisi (2019), o acesso a dados e informações governamentais apresentou crescimento nos últimos anos, juntamente com o aumento de iniciativas de organizações que apoiam a disponibilização desses para incentivar a transparência pública. Essa afirmação está refletida no resultado da pesquisa TIC Domicílios, que aponta que, em 2023, 73% dos usuários de internet no Brasil com 16 anos ou mais utilizaram serviços de governo eletrônico, representando um crescimento de oito pontos percentuais em relação a 2022 (CGI.br/NIC.br; CeticBR, 2023).

Paulo (2018), aponta que os DGA são elementos de base da informação, primários e estruturantes, permitindo a construção de informações significativas e a geração de novos conhecimentos e inovação para a sociedade. Assim, pressupõe-se que a interação dos usuários com os DGA caracteriza-os como informações, e o reúso dessas informações potencializa a geração de conhecimento. Nessa perspectiva, os dados abertos, apesar da denominação como dados, possuem um contexto, podendo ser considerados também uma informação, sendo inseridos no campo de estudo da CI. Além disso, os processos da CI estão diretamente relacionados ao ciclo dos dados abertos, perpassando sua representação, organização e recuperação.

Segundo Santos, Simionato e Arakaki (2014), boa parte da produção informacional está armazenada em ambiente Web, porém ocorrem dificuldades nos processos de acesso, busca e recuperação de informações que interessam aos estudos da CI. No que se refere à representação da informação, em sistemas naturais ou artificiais, ela corresponde ao uso de códigos para a transmissão eficiente da mensagem. O estudo do processamento e de técnicas aplicadas aos computadores e aos sistemas de programação também estão incorporados nas pesquisas sobre representação, com o propósito de armazenagem, recuperação e disseminação ideal (Borko, 1968).

Sant'Ana (2016) considera que a CI pode prover contribuições nesse cenário, sob um novo enfoque, centrado nos dados. Araújo Júnior e Souza (2018) exemplificam essa afirmação ao pressupor que a representação e a organização da informação contribuem para a plena garantia de acesso às facilidades de armazenamento, processamento e entrega de dados, e que a definição de

um modelo de organização da informação e do conhecimento é determinante para a construção de uma solução de *big data*.

No panorama nacional, existem algumas legislações que se destinam a criar mecanismos para possibilitar o acesso a informações públicas por parte de qualquer cidadão. Entre elas se destacam a Lei Nº 12.527, de 18 de novembro de 2011, conhecida popularmente como Lei de Acesso à Informação (LAI), que regulamenta o direito previsto na Constituição de acesso à informação pública, e o Decreto Nº 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo Federal e instituiu a Infraestrutura Nacional de Dados Abertos (INDA), regulamentada pela Instrução Normativa Nº 4, de 12 de Abril de 2012.

Pensando nas interseções entre os dados abertos e a informação - como objeto de estudo da CI - pretendeu-se refletir sobre as aplicações dos instrumentos da CI no ambiente de dados abertos do governo federal brasileiro.

1.1 Problema e Justificativas

Os objetivos da publicação dos DGA são ambiciosos e suas possibilidades de utilização são extensas. Entretanto, diante de fenômenos como o *big data*, que envolvem um grande e heterogêneo volume de dados, surgem preocupações em relação à representação, ao acesso e à recuperação desses dados, a fim de permitir sua efetiva utilização, entre outros aspectos relacionados ao seu manuseio.

Com a expansão dos DGA, os problemas de representação e recuperação relacionados à grande quantidade e diversidade de dados e informações são perceptíveis, sendo essencial conhecer as abordagens e métodos utilizados para representar e recuperar esses dados (Nascimento Silva, 2023).

Em 2024, após diversas legislações sobre os dados abertos no Brasil, observa-se um esforço em todas as esferas para o cumprimento das legislações que dispõe sobre o acesso à informação e a disponibilização de dados abertos, no entanto, a padronização existente ainda é limitada, refletida principalmente na Política de Dados Abertos, que preconiza recomendações, mas não detalha a organização dos dados e informações. Essa situação resulta em publicações de dados dispersos e fragmentados, dificultando sua recuperação e utilização e impossibilitando a reutilização dos dados em outros contextos. Vale ressaltar que, ao longo do tempo, apesar de somar-se aos padrões do governo, como a Infraestrutura Nacional de Dados Abertos (INDA),

portarias, normativas e resoluções, a política ainda está focada em recomendações e boas práticas para a publicação dos dados (Nascimento Silva, 2018; Nascimento Silva, 2022).

A publicação dos DGA auxilia no cumprimento do objetivo da transparência ativa, proposta na LAI (Brasil, 2011), que se dá quando a administração pública opera mediante a publicação proativa de informação, sem esperar a demanda (Saraceno; Monteiro, 2021). A padronização desse processo garante a interoperabilidade dos dados, que é considerada fundamental para que os princípios da abertura se concretizem, resultando em mais e melhores produtos e serviços (*Open Knowledge* Brasil, s.d.a).

A Política de Dados Abertos do Poder Executivo Federal Brasileiro, instituída pelo Decreto nº 8.777, de 11 de maio de 2016, estabelece normas para a livre utilização das bases de dados e assegura a possibilidade de uso e reúso por parte da sociedade (Brasil, 2016). O Decreto também definiu um prazo para que todos os órgãos do Poder Executivo Federal publicassem seu Plano de Dados Abertos (PDA), um documento essencial para organizar e padronizar a publicação de dados nos órgãos públicos, além de estabelecer a INDA, com caráter gerencial e normativo.

Em 2019, foi publicado o Decreto nº 9.903, de 2019, determinando que a Controladoria Geral da União (CGU), por meio da INDA, gerencie a Política de Dados Abertos do Poder Executivo Federal: "A INDA é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos" (Brasil, 2023, *online*).

Nesse contexto, em relação à representação dos dados e da informação, a política recomenda a utilização do Padrão de Metadados do Governo Eletrônico (e-PMG), criado para estabelecer o conjunto mínimo de elementos para a recuperação e o gerenciamento de informações (Brasil, 2014). Contudo, o padrão foi criado para o governo eletrônico, envolvendo conceitos gerais, e a sua última atualização ocorreu em julho de 2014.

A CGU estabeleceu um plano de ação da INDA, com vigência de 2023 a 2025, prevendo a realização de sete ações para fortalecer a Política de Dados Abertos do Poder Executivo Federal, e a ação de número 6 prevê a definição de um "modelo de catalogação de conjuntos de dados para que os órgãos e entidades disponibilizem metadados que sejam gerenciados pelo poder público e fomentem a sua descoberta para ampliar o reúso de dados pelo governo, pelas

empresas e pelas organizações sem fins lucrativos, a fim de estimular a melhoria da prestação dos serviços públicos e privados e a inovação no país" (CGINDA, 2023, p. 12). Essa iniciativa demonstra a importância de práticas que abordam a representação descritiva dos dados públicos do governo de forma a torná-los acessíveis e reutilizáveis.

Ferreira (2022, p. 2) destaca que "a representação descritiva de recursos informacionais de diferentes tipos e formatos é um aspecto fundamental de investigações empíricas e científicas em Biblioteconomia e Ciência da Informação (BCI)", abordagem que foi utilizada no presente trabalho para responder à seguinte questão de pesquisa: **quais metadados podem ser utilizados para representar os dados abertos do governo federal brasileiro?**

Nesta pesquisa, pautou-se em uma lacuna da Representação do Conhecimento e da Informação na Web, apontada por Lima (2020, p. 1), que é a "grande quantidade de dados não estruturados e a dificuldade cada vez maior de acesso a esse volume de informação e sua recuperação". A autora ainda afirma que, no contexto das constantes inovações tecnológicas da informação e da comunicação, são necessárias estruturas de representação e ferramentas de acesso que provenham tratamento desse novo ambiente e tipo de recurso informacional.

Como contribuições, foram sugeridas recomendações (estrutura, organização e descrição) das informações o do Portal Brasileiro de Dados Abertos, resultando em uma proposta de modelo de metadados para a representação descritiva e, consequentemente, a padronização dos DGA do governo federal brasileiro no Portal Brasileiro de Dados Abertos. Destaca-se a correspondência desses metadados com um padrão de metadados específico para catálogos governamentais, indo ao encontro à Ação 6 do Plano de Ação 2023-2025, da INDA, elaborada a partir do desenvolvimento da pesquisa teórica, metodológica e prática (pragmática).

1.2 Objetivos

Nos próximos tópicos, serão apresentados o objetivo geral e os objetivos específicos que nortearam esta dissertação.

1.2.1 Objetivo geral

O objetivo geral deste trabalho foi propor um modelo de metadados para a representação e a organização da informação no Portal Brasileiro de Dados Abertos.

1.2.2 Objetivos específicos

Para alcançar o objetivo geral proposto, foram traçados os seguintes objetivos específicos:

I. Identificar, na literatura acadêmica brasileira, modelos, métricas, *frameworks* e instrumentos similares que abordem a publicação, a avaliação, o ciclo de vida ou o tratamento de dados abertos ou DGA, a fim de evidenciar abordagens, técnicas e métodos utilizados;
II. Mapear documentos e legislações (leis, decretos, normativas, portarias, resoluções e instrumentos) sobre os dados abertos no governo federal brasileiro;
III. Identificar e analisar as tipologias e os formatos de dados abertos publicados, padrões e metadados utilizados no Portal Brasileiro de Dados Abertos;
IV. Propor uma estrutura de representação descritiva para os dados abertos do governo federal brasileiro no Portal Brasileiro de Dados Abertos com a utilização de metadados.

1.3 Estrutura da dissertação

Esta dissertação está organizada em sete seções, divididas da seguinte forma: na **Seção 1**, apresenta-se uma introdução do tema e uma contextualização do trabalho com a área da CI, apresentando o problema de pesquisa, as justificativas e os objetivos - geral e específicos. Na **Seção 2**, apresentam-se as interseções dos DGA com a CI. Na **Seção 3**, apresenta-se a fundamentação teórica, com revisão de literatura para conceituação da temática de organização e representação da informação. Na **Seção 4**, apresentam-se os procedimentos metodológicos e as técnicas empregadas para a realização do trabalho. Na **Seção 5**, apresentam-se os resultados das três primeiras etapas da pesquisa, correspondentes aos três primeiros objetivos específicos. Na **Seção 6**, apresenta-se o resultado da etapa quatro, que é a proposta de modelo de metadados para a representação dos dados abertos no Portal Brasileiro de Dados Abertos, e a discussão pertinente, correspondendo ao quarto objetivo específico de pesquisa. Por fim, na **Seção 7**, apresentam-se uma síntese do trabalho realizado e as conclusões da pesquisa.

2 DADOS ABERTOS NA CIÊNCIA DA INFORMAÇÃO

Nessa seção, apresenta-se uma discussão sobre os conceitos de dados, informação e conhecimento, para introduzir os DGA, que, apesar da denominação relacionada a dados, estão intrinsecamente ligados com os demais. Em seguida, expõe-se um breve histórico dos DGA e o panorama nacional.

Setzer (1999) afirma que os dados, desde que sejam inteligíveis, sempre são interpretados por alguém como informação, uma vez que os seres humanos estão constantemente em busca de significado e compreensão. Com isso, é importante destacar que pode haver uma confusão entre os conceitos, devido à dificuldade de se estabelecer uma fronteira clara entre as definições, como apontado por Davenport (2002). O autor considera difícil definir informação, afirmando que a distinção entre dados, informação e conhecimento é imprecisa, uma vez que informação é um termo que envolve todos os três, servindo de conexão entre os dados brutos e o conhecimento que se pode eventualmente obter.

2.1 Dados, informação e conhecimento

A conceituação de dados, informação e conhecimento não tem uma definição unívoca e universal na CI. Diversos autores apresentam suas reflexões e ponderações, que podem ser adequadas a contextos variados, conforme apresentado nas definições a seguir.

Setzer (1999) apresentou uma definição para dado e realizou uma caracterização dos conceitos de informação, conhecimento e competência. O autor definiu os dados como uma sequência de símbolos que foram quantificados ou podem ser quantificáveis, sendo, assim, passíveis de armazenamento e processamento por um computador. Para ele, os dados são necessariamente entidades matemáticas e, portanto, puramente sintáticos, implicando que eles podem ser descritos por meio de representações formais e estruturais.

Miranda (1999), após uma revisão de literatura, adota a definição de dado como o "conjunto de registros qualitativos ou quantitativos conhecido que organizado, agrupado, categorizado e padronizado adequadamente transforma-se em informação", e a informação como "dados organizados de modo significativo, sendo subsídio útil à tomada de decisão" (Miranda, 1999, p. 286-287).

A visão apresentada por Souza (2006) é convergente com a definição de Setzer (1999), definindo os dados como sequências de símbolos dotados de significado, os quais podem ser codificados, interpretados e processados por sistemas computacionais, além de serem transmitidos a partir de redes de comunicação.

No que concerne ao termo informação, Setzer (1999) a caracteriza como uma abstração informal, que não pode ser formalizada por meio de uma teoria lógica ou matemática. Ela representa algo significativo para alguém mediante textos, imagens, sons ou animações. Segundo o autor, a informação não pode ser processada diretamente em um computador sem que antes seja convertida em dados.

Para Setzer (1999), a representação da informação poderia, eventualmente, ser feita por meio de dados e, assim, ser armazenada em um computador. No entanto, o autor destaca que o que é armazenado na máquina não é a informação em si, mas sim a sua representação na forma de dados. Essa representação pode ser manipulada pela máquina, mas o seu significado não é alterado, pois este depende de quem está interagindo com a informação.

Uma outra conceituação, que é popular na CI, para informação é aquela descrita por Buckland (1991), que considera o uso que se dá para definir o termo:

a) Informação como processo: nesse sentido, refere-se ao ato de informar;
b) Informação como conhecimento: usado para denotar o que se é percebido na informação como processo;
c) Informação como coisa: atribui-se o termo para objetos, assim como dados para documentos, que são considerados como "informação" no sentido de serem informativos, tendo a qualidade de conhecimento comunicado ou comunicação.

Na informação tratada como processo, Buckland (1991) considera que informação é o ato de informar, no sentido de que quando alguém é informado, aquilo que ele conhece é modificado no processo de comunicação.

Quanto à informação como conhecimento, Buckland (1991) considera que ela é intangível, pois não se pode tocá-la ou medi-la de modo algum. O autor reitera que conhecimento, convicção e opinião são atributos individuais, subjetivos e conceituais, contudo, para comunicá-los, eles precisam ser expressos, descritos ou representados de maneira física, como sinal, texto ou comunicação. Com isso, consequentemente, qualquer expressão, descrição ou representação cairia na definição de informação como coisa.

Em relação à terceira definição, informação como coisa, Buckland (1991) considera que o termo dados, em seu significado de “coisas que podem ser dadas”, é uma significação adequada para informação como coisa que tenha sido processada de alguma forma para uso posterior. Essa perspectiva pode ser apropriada para a abordagem dos DGA na CI.

Adota-se, na presente dissertação, a abordagem da informação como coisa, como proposto por Buckland (1991), sendo algo quantificável, mensurável e considerado como informativo para determinada comunidade.

Setzer (1999) destaca que uma distinção crucial entre dados e informação é que os dados são puramente sintáticos, enquanto a informação contém necessariamente semântica, sendo que os dados são objetivos por natureza, independentemente do usuário, enquanto a informação é objetiva-subjetiva, pois é descrita de maneira objetiva (por meio de textos, imagens etc.), mas o seu significado é subjetivo e depende do usuário.

As autoras Brascher e Café (2008) consideram que para compreender informação, deve-se englobar aspectos no nível semântico (cognitivo) e pragmático (real), incluindo as propriedades relativas tanto ao conteúdo e ao significado como à função social, corroborando a visão apresentada por Setzer (1999).

Para o termo conhecimento, Setzer (1999) elabora que é uma abstração pessoal de algo que foi experimentado por alguém. Dessa forma, esse conceito não pode ser descrito totalmente, senão seria meramente reduzido a dado (se fosse descrito formalmente e não tivesse significado) ou informação (se fosse descrito informalmente e tivesse significado). O autor afirma que o conhecimento não depende apenas de uma interpretação pessoal, como a informação, pois ele necessita de uma vivência do objeto do conhecimento.

Fernandez-Molina aponta o relacionamento e a diferenciação entre os termos: “dados são informações potenciais, que, somente se percebidas pelo receptor, transformam-se em informação, e esta se torna conhecimento no momento em que produz uma modificação na estrutura de conhecimento de um destinatário” (Fernandez-Molina, 1994, p.328, tradução nossa).

A visão de Setzer (1999) em relação à diferenciação dos termos (dado, informação, conhecimento) aponta que os dados que representam uma informação podem ser armazenados em um computador, porém a informação não pode ser processada pela máquina quanto ao seu significado, visto que isso depende de quem a recebe. Já em relação ao conhecimento, este não pode nem ser inserido em um computador por meio de representação, senão seria reduzido a uma

informação. Com isso, a informação está associada à semântica enquanto o conhecimento está associado à pragmática, relacionado a algo existente no mundo real, do qual se tem uma experiência direta.

Conforme apresentado pelos teóricos, observa-se que, em muitos casos, a linha de distinção entre os termos é deveras tênue e, muitas vezes, sequer existe. Com isso, é preciso analisar cuidadosamente o contexto para trabalhar com um conceito ou transitar entre eles. Em relação à presente pesquisa, acredita-se que o conceito dos DGA transite entre as definições de dados e informação selecionadas nessa fundamentação teórica. Assim, em relação aos dados abertos, adota-se a visão de que eles são a representação da informação e na medida em que são disponibilizados, interpretados, usados e reutilizados, convertem-se em informação.

2.2 *Open Government Data* (Dados Governamentais Abertos)

O termo *open government data* inclui a junção dos conceitos de dados abertos (*open data*) e governo aberto (*open government*). No Brasil, o termo possui variações e é traduzido como: "dados abertos governamentais", "dados abertos do governo" e "dados governamentais abertos", sendo esta última a nomenclatura utilizada pelas primeiras pesquisas sobre a temática no Brasil, como Albano (2014), e pelo Plano de ação 2023-2025, da INDA, documento mais recente da CGU (2023), valendo-se também nesta dissertação.

O *Open Knowledge* Brasil apresenta uma clássica definição para o termo 'dados abertos', definindo-os como aqueles que "podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras" (*Open Knowledge Brasil*, s.d.b, *online*). Além disso, para fazer jus a essa classificação, segundo a organização, os dados devem seguir os seguintes requisitos:

- **Disponibilidade de acesso**: os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente passíveis de serem baixados pela internet. Os dados devem, também, estar disponíveis de uma forma conveniente e modificável.
- **Reutilização e Redistribuição**: os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados.

- **Participação Universal**: todos devem ser capazes de usar, reutilizar e redistribuir. Não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos [...] (*Open Knowledge* Brasil, s.d.b, *online,* grifo nosso).

Segundo a *Open Knowledge Foundation* (s.d, *online*), o governo é uma das organizações que mais coletam dados e, além da quantidade, a centralidade e a publicidade, por lei, de sua maioria tornam o governo uma fonte muito atrativa de dados abertos. A abertura desses dados poderia favorecer a transparência, a participação, o engajamento e a agregação de valor social e comercial.

Agune, Filho e Bolliger (2010) conceituam governo aberto como a disponibilização, pela Internet, de informações e dados governamentais de domínio público, permitindo seu uso livre pela sociedade, propiciando o acesso a dados primários para sua combinação, cruzamento e geração de novas informações.

Segundo Freitas (2019), o conceito de governo aberto representa uma expressão de gestão apoiada pela tecnologia, visando a uma maior aproximação com o cidadão. Isso é possível ao viabilizar a coautoria dos cidadãos nas decisões governamentais, por meio da participação eletrônica, e garantindo acesso às informações públicas e às suas bases de dados, incluindo aquelas disponíveis em formato aberto.

Acompanhando a tendência de expressiva presença cidadã nas redes, surgiram as iniciativas de governo aberto, que abrangem não somente o fornecimento de dados e informações, mas também o envolvimento efetivo da sociedade nas decisões e ações governamentais (Berberian; Mello; Camargo, 2014)

O ex-presidente dos Estados Unidos da América (EUA), Barack Obama, foi pioneiro em iniciativas relacionadas a governo aberto ao estabelecer, logo no início da sua gestão, três pilares fundamentais para as ações governamentais: transparência, participação e colaboração (Pinho; Silva, 2019), com a publicação do documento *Memorandum on Transparency and Open Government*. Em 2009, foram definidas diretrizes do governo dos EUA para a sua gestão, relacionadas ao acesso aos dados do governo americano, abrangendo desde a gestão interna da informação até a publicação dos dados do governo para o público (Ribeiro; Almeida, 2011).

O termo "open *government data" foi* criado em 2007, em Sebastopol, na Califórnia, durante um encontro com foco na abertura de informações governamentais, que reuniu pesquisadores, representantes de organizações da sociedade civil e ativistas norte-americanos.

Neste encontro, foram definidos oito princípios para dados governamentais, que foram expandidos para outros tipos de dados abertos, estabelecendo que os dados devem ser:

1. Completos: todos os dados devem estar completos e disponíveis;
2. Primários: os dados são apresentados tal como coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação;
3. Atualizados: os dados devem ser disponibilizados tão rapidamente quanto necessário à preservação do seu valor;
4. Acessíveis: os dados devem estar acessíveis: devem ser disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
5. Processáveis por máquina: os dados devem ser processáveis por máquina: devem estar razoavelmente estruturados de modo a possibilitar um processamento automatizado;
6. Acesso não discriminatório: os dados devem estar disponíveis para todos, sem exigência de requerimento ou cadastro;
7. Não proprietários: os dados devem estar disponíveis em um formato sobre o qual nenhuma entidade detenha controle exclusivo;
8. Livres de licença: os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas à privacidade, à segurança e aos privilégios de acesso são permitidas (*Open Knowledge* Brasil, 2015, *online*).

Possamai (2016) aponta que os DGA permitem acesso facilitado às informações e aos dados públicos, que são necessários ao controle e à participação social, além de viabilizarem a integração e a interoperabilidade desses dados, de modo a possibilitar o reúso dos dados e desenvolvimento de novas aplicações e serviços.

Ribeiro e Almeida (2011) definem os DGA como o esforço para a publicação e a disseminação das informações do setor público na *internet*, o que favorece a reutilização, o compartilhamento e a integração destes dados, sendo englobados todos os aspectos da atividade governamental, utilizando, para isso, processamento automatizado em computadores, de acordo com suas necessidade de informação.

Os DGA são regidos por três leis, segundo Eaves (2009): a) se o dado não pode ser encontrado e indexado na Web, ele não existe; b) se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e c) se algum dispositivo legal não permitir sua replicação, ele não é útil. Esses dados influenciam as políticas públicas de três formas:

- Inclusão: fornecer dados em formatos padronizados abertos e acessíveis permite que qualquer cidadão utilize qualquer ferramenta de *software* para adaptá-los às suas necessidades.
- Transparência: informações do setor público abertas e acessíveis melhoram a transparência, pois as partes interessadas podem usá-las da maneira mais adequada ao seu propósito.
- Responsabilidade: os conjuntos apropriados de dados abertos, devidamente associados, podem oferecer vários pontos de vista sobre o desempenho dos governo no cumprimento de suas metas em políticas públicas (Brasil, s.d.h, *online*).

Como complemento, Paulo (2018) afirma que os DGA são uma forma de divulgação proativa de informação, disponibilizada *online*, e que "sua publicação na esfera pública deve estar inserida em uma estratégia informacional que contemple o planejamento, o controle e a mensuração de resultados mediante uma compreensão clara de sua relevância e valor associados à sociedade, ao desenvolvimento da cidadania e à própria democracia" (Paulo, 2018, p. 12). Possamai (2016) e Cunha (2020) fazem coro a essa perspectiva ao considerarem os DGA como pilares da democracia.

Os DGA estão relacionados ao conceito de dados, quando disponibilizados de forma primária, em conjuntos de dados, e também à informação, quando esses conjuntos são contextualizados e apresentados em portais e painéis. Com isso, é na CI que os DGA estão fundamentados, assim como as técnicas de organização, representação e recuperação que irão subsidiar a manipulação destes dados e informações para seu efetivo reúso. Destaca-se que a temática dos DGA ainda é recente e essa associação com a CI nem sempre é visualizada nas pesquisas que envolvem os dados abertos e os DGA.

3 FUNDAMENTAÇÃO TEÓRICA

Na presente seção, serão apresentados conceitos e reflexões da literatura que embasaram as discussões propostas nesta dissertação, iniciando pelas marcações conceituais de organização, representação e recuperação da informação no contexto da CI, aprofundando a temática de representação da informação, com a apresentação dos instrumentos utilizados para a proposição apresentada nesta pesquisa, e finalizando com a governança de dados.

3.1 Organização, representação e recuperação da informação

A organização e o tratamento da informação e do conhecimento fazem parte dos fluxos informacionais que regem os estudos da CI e, segundo Lima (2020), essa área compreende os estudos e instrumentos utilizados na organização de recursos informacionais de qualquer natureza, visando propiciar o atendimento às necessidades de informação de uma comunidade de usuários. A autora também afirma que a organização da informação infere um processo subdividido em dois subprocessos complementares: o armazenamento e a recuperação (Lima, 2020). Considerando essa perspectiva e o contexto informacional altamente tecnológico vivenciado atualmente, os DGA representam consideráveis recursos informacionais a serem organizados e tratados na perspectiva da CI.

Souza e Almeida (2009) acreditam que a CI se preocupa com processos de representação do conhecimento e de seus registros associados (representação da informação), e é do cunho destes processos que a informação passe por abstrações, modelagens e representações para ser organizada, transmitida, codificada e consumida.

Brascher e Café (2008, p. 5) afirmam que o "objetivo do processo de organização da informação é possibilitar o acesso ao conhecimento contido na informação". Essa visão é consonante com as afirmativas de Ortega (2013), que define a OI como um conjunto de procedimentos realizados sobre documentos para propiciar seu uso por públicos específicos, de acordo com necessidades de informação de ordem científica, educacional, profissional, estética, de entretenimento e utilitária. Sendo um dos objetivos dos DGA propiciar acesso aos dados públicos para os cidadãos, acredita-se que os processos de OI precisam necessariamente estar envolvidos na representação, publicação e disponibilização dos DGA para que esse objetivo seja cumprido.

Assim como a conceituação de informação e conhecimento, ocorrem discussões, convergências e divergências na CI sobre a aplicação dos termos Organização do Conhecimento (OC), Organização da Informação (OI), Representação do Conhecimento e Representação da Informação (Brascher e Café, 2008). Para essas autoras, existem dois tipos de processos de organização, sendo um aplicável às ocorrências individuais de objetos informacionais - o processo de OI, e outro, aplicável a unidades do pensamento (conceitos) - processo de OC (Brascher e Café, 2008).

Adere-se, nesta pesquisa, à linha da OI, que compreende a organização de um conjunto de objetos informacionais (os DGA, neste caso, concreto) com o objetivo de arranjá-los sistematicamente em coleções (Brascher e Café, 2008). Já em relação à OC, esta visa a "construção de modelos de mundo que se constituem em abstrações da realidade" (Brascher e Café, 2008, p. 6).

O OI resulta na representação da informação, sendo o conjunto de atributos que representa determinado objeto informacional, obtida por meio de descrição física (representação descritiva) e de conteúdo (representação temática), e a OC resulta na representação do conhecimento, constituindo-se numa estrutura conceitual que representa modelo de mundo (Brascher e Café, 2008).

Os autores Mello Filho e Araújo Júnior (2021) afirmam que é essencial alinhar as necessidades informacionais do cidadão com a política de disponibilização de dados pela administração pública, a partir de uma estrutura de organização de informação. Nesse contexto, Guedes e Araújo Júnior (2014) consideram a recuperação de informação como a função mais significativa em um sistema de comunicação. Os autores Baeza-Yates e Ribeiro-Neto trazem uma clássica definição:

> A Recuperação de Informação trata de representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas Web, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse (Baeza-Yates; Ribeiro Neto, 2013, p.1).

Essa visão sobre a recuperação de informação é concordante com a apresentada por Ferneda (2012), que definiu o processo de recuperação de informação como aquele que compreende a identificação em um conjunto de documentos, os que atendem à necessidade de informação de um usuário.

Esse processo ocorre no âmbito dos sistemas de recuperação de informação (SRI), definidos por Cesarino (1985) como um conjunto de operações sequenciais, que são realizadas com o objetivo de encontrar, dentre todas as informações disponíveis, aquelas relevantes ao usuário. Ferneda (2012) complementa essa definição indicando que os SRI devem representar o conteúdo dos documentos, apresentando-os ao usuário de forma que ele possa realizar uma rápida seleção dos itens que satisfazem, em parte, ou na totalidade, suas necessidades de informação.

A recuperação de informação pode ser abordada sobre dois pontos de vista, sendo (i) o do computador, focado na construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento, ou (ii) focado no usuário, estudando seu comportamento a fim de entender suas principais necessidades e determinar como esse entendimento afeta a organização e a operação do sistema de recuperação (Baeza-Yates; Ribeiro-Neto, 2013).

Nas subseções seguintes, aprofundam-se mais as discussões sobre representação da informação, que são o foco principal da proposta desenvolvida.

3.2 Representação da informação

A representação da informação é um conceito central na OI e, segundo Brascher e Café (2008), envolve a descrição física e de conteúdo dos objetos informacionais, sendo "entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico" (Brascher e Café, 2008, p. 5). Esta descrição possibilita uma comunicação eficaz entre o usuário e o documento, conforme apresentam Maimone, Silveira e Tálamo (2011) ao afirmarem que tanto a representação temática quanto a descritiva utilizam-se de uma linguagem própria para viabilizar esse objetivo.

Hjorland (2008) considera que, além da descrição de conteúdos, a organização dessas descrições é necessária para que elas possam ser acessadas e recuperadas, atendendo às necessidades informacionais dos usuários e Lourenço (2016), corrobora ao afirmar que, historicamente, a representação da informação, a partir do registro de informações descritivas e temáticas dos documentos, tem sido essencial para o controle bibliográfico e a recuperação de documentos desde o final do séc. XIX. Nos contextos da Biblioteconomia e da CI, essa representação é realizada por meio dos processos de catalogação, indexação e classificação em um Sistema de Recuperação da Informação (SRI) (Lima, 2020).

Um dos processos para a representação da informação é a catalogação, que, segundo Ortega (2013), é um termo abrangente, que envolve os modelos de registros de bases de dados e aspectos teóricos e técnicos da representação descritiva. Já segundo Maimone, Silveira e Tálamo (2011), a catalogação ganhou duas designações, sendo a parte da catalogação descritiva nomeada como Representação Descritiva e a catalogação de assunto, designada como Representação Temática. Os autores apresentam, ainda, a conceituação de ambos:

> A primeira representa as características específicas do documento, denominada descrição bibliográfica, que permite a individualização do documento. Ela também define e padroniza os pontos de acesso, responsáveis pela busca e recuperação da informação, assim como pela reunião de documentos semelhantes, por exemplo, todas as obras de um determinado autor ou de uma série específica. A segunda detém-se na representação dos assuntos dos documentos a fim de aproximá-los, tornando mais fácil a recuperação de materiais relevantes que dizem respeito a temas semelhantes (Maimone; Silveira; Tálamo, 2011, p. 28).

Brascher e Café (2008) alertam que para que os objetivos da organização da informação sejam alcançados, é necessário realizar tanto a descrição física quanto a descrição de conteúdo dos objetos informacionais.

Segundo Mey (1995), a catalogação consiste na representação de um item para simplificar a sua busca. Em uma definição mais completa, a autora conceitua a catalogação como: “O estudo, preparação, organização de mensagens codificadas, com base em itens existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir interseção entre as mensagens contidas nos itens e as mensagens internas dos usuários” (Mey, 1995, p. 5). A autora apresenta um quadro com as características bases da catalogação:

Quadro 1 - Caracterização da catalogação

Características da catalogação	
Integridade	Fidelidade e honestidade na representação, transmitindo informações passíveis de verificação
Clareza	O código utilizado deve ser compreensível aos usuários
Precisão	No código utilizado, cada informação só pode representar um único dado ou conceito, sem dar margem à confusão entre as informações
Lógica	As informações devem ser organizadas de modo lógico
Consistência	A mesma solução deve ser sempre usada para informações semelhantes

Fonte: adaptado de Mey (1995).

A catalogação é descrita por Alves (2010, p. 12) como “um processo de representação informacional que garante a identificação única do recurso informacional para fins de recuperação e, portanto, essencial para proporcionar a recuperação eficiente dos recursos nos variados tipos de sistemas de informação”. Segundo Lima e Álvares (2012), na CI, a representação está relacionada com as formas de simbolizar a informação e o conhecimento, e sua principal característica é a substituição do texto do documento por uma descrição abreviada.

Carvalho (2022, p. 42) elabora que a representação de um documento “inclui os elementos descritivos que o identificam e o caracterizam em um acervo, assim como os elementos indicativos de seu conteúdo informativo, os assuntos por ele tratados. Tem como objetivo identificar cada documento de um determinado acervo, tornando-o "visível" para os usuários de um sistema de informação.”

Alvarenga (2003) considera que os processos de representação podem ocorrer em momentos distintos, sendo na produção dos registros de conhecimento, na organização dos sistemas de informações documentais ou no acesso às informações por parte dos usuários. Aqui, aborda-se o segundo momento, realizando proposições para a representação e disponibilização dos DGA na web.

Infere-se que a catalogação é uma definição abrangente de processos para construir a representação de quaisquer objetos informacionais, de modo a torná-los acessíveis a um determinado público, possuindo duas dimensões complementares, sendo uma dimensão

descritiva, preocupando-se com a forma e as características físicas, e outra temática, ocupando-se dos conteúdos.

Yamane e Castro (2018) consideram que a representação descritiva abarca as formas de descrição e de representação de recursos informacionais em ambientes informacionais e utilizam códigos de catalogação no processo de representação para realizar a codificação dos dados bibliográficos do recurso informacional. Como exemplo, pode-se citar o *Machine Readable Cataloging* (MARC), o Código de Catalogação Anglo-americano (AACR2), o *Resource Description and Access* (RDA) e o *International Standard Bibliographic Description* (ISBD), amplamente utilizados para descrição de recursos bibliográficos.

Tomoyose (2021) considerou a representação da informação como central na abordagem da disponibilização e do compartilhamento de quaisquer fontes e recursos informacionais, como é o caso dos DGA, por isso, considera-se de extrema relevância o estudo da representação desses recursos. Apesar da representação descritiva e temática serem processos relacionados, especificou-se com mais detalhes os elementos que envolvem a primeira, por estar intrinsecamente relacionada à temática de pesquisa, e utilizou-se o termo representação descritiva nesta dissertação para referir-se ao tratamento descritivo do material, como apontado por Fujita *et al* (2009).

No contexto dos DGA, sendo uma de suas premissas a disponibilização online, preocupa-se também com particularidades da representação da informação na Web, que, desde sua criação, acarreta discussões sobre a organização dos recursos informacionais, como apontado por Grácio (2002), que alertou para a necessidade de ferramentas para o gerenciamento, a organização e o armazenamento dos recursos disponíveis em meio eletrônico devido ao célere desenvolvimento da Web e dos recursos disponíveis na Internet, definindo recurso como toda informação que pode ser acessada e recuperada, independentemente do seu formato.

Zafalon (2012) argumenta que ambientes digitais necessitam de padrões para a representação de documentos eletrônicos, garantindo sua recuperação e seu acesso. A autora também defende que a representação de recursos informacionais é vital para facilitar a comunicação entre unidades de informação e usuários, envolvendo a transmissão de documentos reunidos por suas características de conteúdo e expressão de forma individualizada.

Tomoyose (2021) acredita que é papel do cientista da informação tentar solucionar a desordem da Web intervindo com uma aplicação apropriada da representação, especificamente utilizando-se de metadados, cuja definição será apresentada na subseção seguinte.

3.3 Metadados e padrões de metadados

Na definição mais popular, metadados são descritos como ‘dados sobre dados’. Tratando-se de representação da informação, Lima (2020, p. 80) afirma que “os metadados atuam como referenciais ao item informacional representado, e como intermediários entre o objeto representado e o usuário”.

Castro (2020) considera que o objetivo e a função dos metadados têm seus fundamentos enraizados nos princípios da catalogação descritiva, entre os quais estão garantir a padronização dos recursos informacionais com a utilização de esquemas e regra internacionais, facilitando a identificação, a busca, a localização, a recuperação, a preservação, o uso e o reúso dos recursos informacionais.

Os metadados são definidos por Grácio (2002, p. 21) como “um conjunto de dados chamados de elementos, cujo número é variável de acordo com o padrão, e que descreve o conteúdo de um recurso, possibilitando a um usuário ou a um mecanismo de busca acessar e recuperar esse recurso”. O conjunto de elementos que descrevem o recurso pode ter um número de elementos variável de acordo com o padrão de metadados que utiliza e deve conter dois tipos de informações: i) descritivas, referentes às características explícitas do recurso, como título, data etc e ii) temáticas, de conteúdo intelectual, expressando o conteúdo do recurso como palavras-chave (Grácio, 2002). O autor também fez um compilado com definições encontradas na literatura, apresentadas no Quadro 2:

Quadro 2 - Definições do termo ‘metadados’

Autor	Definição
Madsen, Foog e Ruggles (1994) / Klensin (1995) / Garcia (1999)	“dados sobre dados”
Madsen, Foog e Ruggles (1994)	“dados que descrevem elementos do recurso e também descrevem conjunto de dados e sistemas de gerenciamento de Banco de Dados”
Gilliland-Swetlad (1998)	“dados sobre indexação e catalogação criados para ordenar e tornar mais acessível a informação”
Gill (1998)	“dados que servem para descrever grupos de dados que poderíamos chamar de objetos informatizados” ou “descrições estruturadas de um objeto informatizado”
Ortiz-Repiso Jimenez (1999)	“conjunto de dados usados para descrever e representar um objeto de informação”
Barreto (1999)	“documentação que descreve o dado armazenado”
Vendrusculo e Melo (2000)	“catalogação do dado ou descrição do recurso”
Takahashi (2000)	“dados que descrevem outros dados”
Rossetto (2002)	“dados acerca de dados que estão no espaço digital”

Fonte: adaptado de Grácio (2002, p. 22-23).

Para Alves (2010), os metadados são atributos que representam um recurso informacional e o consideram como elemento essencial no processo de tratamento descritivo da informação, representando a interseção entre aspectos tecnológicos e de representação requeridos aos novos tipos de recursos e ambientes informacionais, resultando em uma recuperação mais eficiente em ambientes digitais e promovendo o intercâmbio e a interoperabilidade entre sistemas.

Considerando a catalogação de recursos informacionais, a definição dos metadados objetiva a escolha dos elementos necessários para descrever diferentes tipos de recursos com a finalidade de definir elementos descritivos que irão compor o registro que representa cada item do acervo de determinada instituição (Santos; Simionato; Arakaki, 2014). Os autores ainda ressaltam que, em ambiente digital, os metadados são influentes no acesso e na recuperação da informação, sendo utilizados como elementos de busca que facilitam o acesso aos dados descritivos e a localização de recursos na web.

Para Santos, Simionato e Arakaki (2014), os metadados evidenciam os diversos aspectos do recurso que descrevem: estrutura, conteúdo, qualidade, contexto, origem, propriedade e condição; além de auxiliarem na organização, favorecerem a interatividade, validarem identificações, assegurarem a preservação e otimizarem o fluxo informacional, resultando em uma melhora no acesso aos dados e na localização dos recursos informacionais.

Para Lourenço (2016), os metadados são a chave para a resolução dos problemas de produção e recuperação de informações da web, visão consonante com a de Tomoyose (2021), que acredita que eles são centrais na organização de dados publicados, atribuindo a eles a responsabilidade da padronização e recuperação desses dados. Carvalho (2022) tem um raciocínio similar ao considerar que somente a representação adequada e padronizada de recursos informacionais com a utilização de metadados pode garantir maior precisão na recuperação e no tratamento da informação.

Arakaki (2019) propôs um novo conjunto de definições e hierarquia para as diversas tipologias de metadados existentes, baseado em uma revisão de literatura, conforme apresentado no Quadro 3:

Quadro 3 - Definição dos tipos de metadados

Tipo	Definição
Metadados administrativos	Metadados administrativos são usados para gerenciar e administrar coleções e recursos informacionais e para auxiliar na tomada de decisão e manutenção dos registros e recursos informacionais. Fornecem informações sobre a origem e a manutenção de um objeto.
Metadados de autenticação	Metadados de autenticação são informações que possibilitam a identificação, integridade e legitimidade de um recurso informacional. Exemplos consistem em: código de identificação ou verificação, assinatura digital, entre outros (Greenberg, 2001).
Metadados de preservação	Metadados de preservação estão relacionados com informações de preservação e conservação dos recursos informacionais.
Metadados de proveniência	Metadados de proveniência estão relacionados às informações de procedência, fornecendo dados sobre entidades, criação e modificações e seus relacionamentos (Pomerantz, 2015).
Metadados técnicos	Metadados técnicos estão relacionados a como um sistema funciona, fornecendo informações do sistema ou do recurso.
Meta-metadata	Meta-metadata corresponde a informações sobre o registro criado, ou informações da criação de um conjunto de dados.
Metadados descritivos	Metadados descritivos descrevem as características identificadoras e os contextos intelectuais dos recursos de informação para fins de descoberta, identificação, seleção, aquisição, contexto e

	compreensão (Joudrey; Taylor, 2018).
Metadados de direitos	Metadados de direitos estão relacionados às informações sobre propriedade e aos direitos autorais.
Metadados de acesso e uso	Metadados de acesso e uso são informações sobre como um recurso informacional foi acessado e usado, como restrições de circulação e acesso, registros de exposições, entre outros.
Metadados estruturais	Metadados estruturais estão relacionados à composição e organização do recurso informacional.
Markup languages	*Markup languages* integram metadados e sinalizações para outros recursos estruturais ou semânticos (Riley, 2017).

Fonte: Arakaki (2019, p. 80-81)

Os metadados delineiam diversos aspectos do recurso que descrevem, abrangendo sua estrutura, seu conteúdo, sua qualidade, seu contexto, sua origem, sua propriedade e sua condição. Além disso, eles representam um papel central na organização, facilitam a interação, validam identificações e garantem a preservação, otimizando o fluxo informacional e aprimorando o acesso aos dados e à localização dos recursos informacionais. Para cumprirem essas funções e auxiliarem na interoperabilidade, tendo aceitação internacional, os metadados devem possuir uma padronização (Santos; Simionato; Arakaki, 2014).

Tomoyose (2021) complementa essa perspectiva afirmando que a representação externa é fruto de um processo cognitivo, ocorrido em cada indivíduo, ressaltando a importância da utilização de diretrizes e padrões para uniformizar as representações, obtendo consistência na organização da informação e propiciando sua recuperação.

Alves (2010) afirma que os metadados geraram diversos padrões para a representação das informações no meio digital, vinculando seu desenvolvimento e sua aplicação aos métodos de organização, tratamento e representação informacional. Na escolha dessa padronização, deve-se considerar não apenas as necessidades institucionais, mas, primariamente, as necessidades de informação dos usuários (Santos; Simionato; Arakaki, 2014). Carvalho (2022) complementa que a utilização de padrões de metadados possibilita a troca de informações entre instituições, dispensando o retrabalho de descrição dos dados, e que uma escolha adequada permite a interoperabilidade entre sistemas com maior eficiência e rapidez.

Segundo Tomoyose (2021), com a utilização de metadados, é possível a identificação, a preservação, o acesso e a recuperação da informação, sendo necessário adotar-se vocabulários

para a padronização e uniformização da aplicação desses metadados, garantido a efetividade dos benefícios. Os vocabulários voltados à padronização dos metadados são, no âmbito da CI, conhecidos como padrões de metadados.

Em relação aos padrões de metadados, eles são descritos por Barreto (1999) como:

> Um conjunto de elementos descritores organizados segundo um modelo de dados que é empregado para descrever objetos do domínio de informação do usuário. Um modelo de dados, por sua vez, pode ser definido como um conjunto de conceitos e regras de composição associadas que são usadas para descrever tipos e relacionamentos entre as instâncias desses tipos. No contexto relacional, o modelo de dados básico consiste de tabelas, registros e atributos. No contexto da orientação a objetos, o modelo de dados básico consiste de classes e objetos. Em ambos os modelos, regras de composição distintas se aplicam. No modelo orientado a objetos, essas regras se referem à herança entre classes, polimorfismo, domínio de atributos e agregação. No modelo relacional, essas regras especificam, por exemplo, que um registro deve conter um número qualquer de atributos e que cada atributo deve ser definido em termos de um tipo primitivo de dado (Barreto, 1999, p. 53).

Em consonância com Barreto (1999), Alves (2010, p. 116-117) afirma que "os padrões de metadados são compostos pelo conjunto de elementos descritivos, os metadados codificados que descrevem um recurso informacional. [...] Os padrões de metadados consolidam em uma estrutura automatizada os metadados presentes nos esquemas de codificação manual do domínio bibliográfico".

Grácio (2002, p. 36) considera que o primeiro padrão de metadados foi o MARC (*Machine Readable Cataloging*), criado nos Estados Unidos, na década de 1960, "com o objetivo de possibilitar a troca de registros bibliográficos e catalográficos entre bibliotecas, com o uso de computadores". O autor afirma que esse tipo de padrão possui um conjunto de elementos complexos e rígidos, que demandam especialistas para a sua descrição, diferentemente dos padrões criados para a descrição de recursos disponíveis na web, que têm a característica de possuir um conjunto de elementos mais simples, em menor quantidade e mais flexíveis, facilitando a descrição pelo próprio autor do recurso ou por administradores de sites (Grácio, 2002).

No contexto desta pesquisa, por meio de pesquisa documental, partindo de publicações disponibilizadas no Portal Brasileiro de Dados Abertos, foram identificados três padrões de metadados relacionados aos dados abertos do governo federal: o Padrão de Metadados do Governo Eletrônico (e-PMG) 1.1, de 2014, o padrão da Cartilha Técnica para Publicação de

Dados Abertos e o recente DCAT-BR, que foram apresentados na seção 5, juntamente com outras iniciativas de dados abertos.

O e-PMG foi criado com base no *Dublin Core* (DC) e o DCAT-BR foi baseado no *Data Catalog Vocabulary* (DCAT) e, por isso, esses dois padrões base serão apresentados nas próximas subseções, pois foram utilizados como base para a proposta dessa dissertação, assim como os vocabulários *Friend Of Friends*, *vCard* e *Adms*, que são complementares ao DCAT.

É importante destacar que existe uma similaridade entre o DC e o DCAT que possibilita a interação entre eles, de forma a apresentar uma modelagem mais completa e detalhista: a utilização do *Resource Description Framework* (RDF), que é um modelo orientador para o processo de descrição dos recursos no contexto do *linked data* (dados ligados/conectados) e permite que a descrição seja feita com a utilização de uma base estrutural única em diferentes domínios, favorecendo o reúso de recursos e dos metadados que os descrevem, promovendo interoperabilidade (Jesus; Santarém Segundo, 2023). A especificação RDF é recomendada pela estrutura de governo eletrônico (e-PING - Padrões para Interoperabilidade de Governo Eletrônico) e pelo *World Wide Web Consortium* (W3C) e foi considerada ao se utilizar padrões que a incluem, como o DC e o DCAT. A seguir, apresentam-se as especificações do DC, do e-PMG, do Vocabulário Controlado de Governo Eletrônico (para definição de categorias de assuntos de temas do governo) e do DCAT.

3.3.1 *Dublin Core* (DC)

O Dublin Core (DC) é um conjunto de 15 elementos genéricos de metadados criados para descrever recursos na *Web* (Alcaraz Martínez, 2023). As primeiras conjecturas sobre o DC foram apresentadas em um *workshop* realizado em Dublin, Ohio e EUA, convocado pela *Online Computer Library Center* (OCLC), em 1995 (Alcaraz Martínez, 2023). O principal resultado deste *workshop* foi a criação de um conjunto com 13 elementos (*Subject, Title, Author, Publisher, OtherAgent, Date, ObjectType, Form, Identifier, Relation, Source, Language e Coverage*), chamado de "*Dublin Core Metadata Element Set*" (Grácio, 2002).

Em 1996, ocorreu a ampliação do DC de 13 para 15 elementos, adicionando os elementos *Description* e *Rights Management*, resultando na versão 1.0 do padrão DC (Grácio, 2002). Em 2002, o *Dublin Core Metadata Initiative* (DCMI) assumiu o papel de agência de padrões,

mantendo sua documentação atualizada para os Termos de Metadados do DCMI (Alcaraz Martínez, 2023).

Baseando-se nos 15 elementos de metadados e para se adaptar à evolução da Web em direção à Web semântica, a DCMI continuou a desenvolver outros padrões com base no modelo de dados RDF (*Resource Description Framework*) proposto pelo W3C (Alcaraz Martínez, 2023).

O DC é baseado em quatro princípios fundamentais, segundo Alcaraz Martínez (2023):

- Simplicidade de criação e manutenção;
- Semântica universal;
- Internacionalização;
- Extensibilidade;

A especificação do DC possui dois níveis: o simples e o qualificado. Todos os elementos são opcionais e reutilizáveis e a ordem de aparição também é opcional (Alcaraz Martínez, 2023). Os elementos são apresentados no Quadro 4:

Quadro 4 - Elementos do DC

Elemento	Descrição	Tipo de informação
title (título)	Nome do recurso	Conteúdo
subject (assunto)	Assunto de que se trata o recurso	
relation (relação)	Apresenta a relação entre recursos	
source (fonte)	Recurso de qual o presente recurso é derivado	
coverage (cobertura)	Área de abrangência do recurso	
type (tipo)	Categoria ou natureza do recurso	
description (descrição)	Descrição sobre o conteúdo do recurso	
creator (autor)	Responsável pela criação do recurso	Propriedade intelectual
contributor (contribuidor)	Entidades que contribuíram com a criação do recurso	
publisher (editor)	Responsável pela disponibilização do recurso	
rights (direitos)	Informações sobre os direitos de uso ou acesso ao recurso	

date (data)	Data associada ao evento no ciclo de vida do recurso, como data de criação, publicação, modificação	Instância
format (formato)	Formato físico ou digital do recurso	
identifier (identificador)	Identificador único do recurso	
language (idioma)	Idioma do conteúdo do recurso	

Fonte: adaptado de Alcaraz Martínez (2023).

Cada um dos elementos é definido por um conjunto de 10 atributos padrões da ISO/IEC 11179 para a descrição dos elementos de dados, sendo eles: nome, identificador, definição, comentário, versão, registro de autoridade, língua, obrigação, tipo de dado e máxima ocorrência, sendo os seis últimos comuns para todos os elementos do DC (Grácio, 2002).

Pavão, Costa, Ferreira e Horowitz (2015) consideram a DCMI como a iniciativa mundial mais conhecida em relação aos esquemas de metadados para a descrição de objetos digitais, principalmente textuais e Arakaki, Alves e Santos (2018) apontam que o DC tornou-se um padrão consolidado ao longo dos 20 anos de sua criação, possibilitando diversas possibilidades de uso. A seguir, trata-se do e-PMG, criado com base no DC.

3.3.2 Padrão de Metadados do Governo Eletrônico (e-PMG)

O Padrão de Metadados do Governo Eletrônico (e-PMG) consiste em um documento de 41 páginas, atualizado pela última vez em 2014, que "visa assegurar às pessoas que pesquisam as informações do governo brasileiro na Web o acesso rápido e eficiente à descrição dos recursos" (Brasil, 2014, p. 4).

O documento estabelece um conjunto mínimo de elementos com os dados necessários para a recuperação e o gerenciamento de informações do governo. Os motivadores para a criação desse padrão foram a modernização do governo e o facilitamento de gerência e recuperação de informações com a utilização de metadados.

Esse padrão foi desenvolvido com base no Padrão *Dublin Core* (DC), do *Dublin Core Metadata Initiative* (DCMI), e consiste em 20 elementos: 15 elementos do DC e 5 elementos adicionais (contexto jurídico-administrativo, destinação, destinatário, localização, preservação) considerados necessários para o contexto do governo eletrônico brasileiro. O e-PMG se difere do

DC em alguns aspectos, como na possibilidade de ser utilizado para descrever tanto recursos eletrônicos quanto não eletrônicos e por possuir alguns qualificadores adicionais e alterações nos critérios de obrigatoriedade ou não do uso de elementos e qualificadores (Brasil, 2014).

Segundo o manual do e-PMG 1.1 de 2014, os próprios órgãos do governo devem avaliar seus objetivos e público-alvo para decidir quais recursos devem ser descritos com o uso do e-PMG. O manual apresenta-se como uma diretriz, e não uma norma obrigatória, o que interpõe a continuidade dos desafios em busca da padronização e interoperabilidade. É recomendado que os metadados sejam utilizados o mais cedo possível no ciclo de vida de um recurso, quando ele é criado ou publicado e utilizem serviços profissionais relacionados à gestão da informação para apoiar a criação, a adequação e o controle de qualidade dos metadados e-PMG (Brasil, 2014).

Os metadados e-PMG são flexíveis, podendo ser acrescentados ou atualizados a qualquer momento, porém existem metadados que precisam ser obrigatoriamente atualizados sob o risco de invalidar o esforço no processo de criação e uso. Para cada elemento incluído no padrão, existem dados que são descritos no Quadro 5:

Quadro 5 - Dados dos elementos do e-PMG

Dado	Característica
Nome	Termo atribuído ao elemento em português
Identificador	Termo atribuído ao elemento de acordo com o padrão originário
Definição	Uma descrição do que se trata o elemento
Obrigatoriedade	Define o grau de uso do elemento, podendo ser: ● Obrigatório: este elemento deve obrigatoriamente ter um valor; ● Obrigatório se aplicável: a este elemento deve ser fornecido um valor se o tipo de recurso assim o requerer; ● Opcional: a este elemento pode ser fornecido um valor se o dado estiver disponível e apropriado ao recurso; A obrigatoriedade aplica-se ao elemento e seus qualificadores, quando for o caso.
Objetivo	Fornece a informação essencial sobre a aplicabilidade ao elemento
Comentários	Informação adicional para a compreensão e utilização do elemento ou dos qualificadores do elemento
Não confundir com	Fornece informações comparativas com elementos que eventualmente possam gerar dúvidas ou semelhanças no nome e suscitar uso inadequado
Qualificadores	Usado para tornar o significado de um elemento mais específico, podendo ser utilizado para informação adicional de um recurso

Exemplos	Indica como os elementos podem ser acrescentados para diferentes situações. Os exemplos são fictícios.
Sintaxe de HTML	Visa facilitar a forma de uso dos elementos para páginas da Web
Esquemas	Indicam a fonte que especifica a forma ou informa os valores possíveis para um elemento. Incluem vocabulários controlados, normas brasileiras e internacionais
Mapeamento	Correlação com elementos de outros padrões de metadados, como: e-GMS: Government Metadata Standard AGLS: Australian Government Locator Service GILS: Government Locator Service FGDC: Federal Geographic Data Committee (FGDC) Metadata MARC: Machine Readable Cataloging

Fonte: adaptado de Brasil (2014)

Em relação aos 20 elementos de metadados do e-PMG, eles são apresentados de forma sucinta no Anexo 1. Observa-se que o documento apresenta riqueza de detalhes e exemplificação de utilização, porém, a não obrigatoriedade de aplicação e a falta de atualização do mesmo trazem o risco da obsolescência devido ao desuso. Apresenta-se, a seguir, o Vocabulário Controlado de Governo Eletrônico, ferramenta utilizada para a representação temática no âmbito do governo federal.

3.3.3 Vocabulário Controlado de Governo Eletrônico (VCGE 2.1.0)

O Vocabulário de Controlado Governo Eletrônico (VCGE) é um vocabulário controlado criado para indexar informações, consideradas aqui como documentos, bases de dados, sites etc, no governo federal que tem os objetivos de ser uma interface de comunicação com o cidadão, indexando as informações de forma simples e compreensível e de ser uma ferramenta de gestão para gerenciamento de informações (Brasil, 2023, *online*).

O VCGE surgiu a partir de uma lista de termos chamada Lista de Categorias de Governo (LCG), criada em março 2004, que contemplava assuntos relacionados à atuação do governo e passou a ser conhecida por Lista de Assuntos de Governo (LAG) dois anos depois, sendo focada em taxonomia de navegação (Brasil, 2016*;* Brasil, 2024*, online*).

Em novembro de 2010, surgiu o VCGE, que teve sua difusão e uso restritos até o advento da LAI, em 2011, quando ele ganhou o importante papel de classificar todas as demandas da sociedade pelo Sistemas de Informação ao Cidadão da CGU. O VCGE era mantido por grupos de

voluntários e indicados pelos órgãos, sendo um dos grupos de trabalho da e-PING. Em 2012, iniciou-se um esforço para a reavaliação do VCGE, que teve como resultados um *workshop* e uma série de reuniões com a publicação da versão 2.0, em outubro de 2013 (Brasil, 2016, *online*; Brasil, 2023, *online*).

Na data da pesquisa, realizada em agosto de 2024, a versão mais atualizada, 2.1.0, de 2016, encontrava-se inacessível no endereço eletrônico: https://www.gov.br/governodigital/pt-br/governanca-de-dados/vocabulario-controlado-do-governo-eletronico. Para a apresentação deste instrumento, foi utilizada uma cópia digital, disponibilizada no âmbito do projeto Observatório de Dados Abertos.

O objetivo do VCGE, descrito como um vocabulário controlado para indexar informações (documentos, bases de dados e sites) no governo federal, não é ser um grande vocabulário de governo, contudo, ele tem dois objetivos básicos:

- Ser uma interface de comunicação com o cidadão, indexando informações de forma simples e entendível;
- Ser uma ferramenta de gestão, auxiliando os gestores no gerenciamento de informações (Brasil, 2016).

O VCGE é um vocabulário poli hierárquico, podendo um termo ter mais de um "pai", mas não é multi hierárquico, não existindo hierarquias paralelas, e adota os conceitos de 'termo preferencial', 'termo geral' e 'termo relacionado'. O nível 1 possui 26 termos que são utilizados para a classificação de categorias de conjuntos de dados no Portal Brasileiro de Dados Abertos, até a data da pesquisa em novembro de 2024. Na próxima subseção, apresenta-se o padrão de metadados selecionado para aplicação nesta dissertação, o *Data Catalog Vocabulary* (DCAT-3).

3.3.4 *Data Catalog Vocabulary* (DCAT-3)

Os vocabulários são um aspecto intrínseco das relações entre a CI e o contexto de publicação de dados na Web, sendo que, no âmbito da CI, o conceito de vocabulário se estabelece inicialmente como um produto da OI e da OC, como um instrumento para a padronização das atividades de controle de vocabulário (Jesus; Santarém Segundo, 2023).

O conceito de vocabulário caracteriza-se como um termo abrangente, utilizado para referir-se a diferentes tipos de Sistemas de Organização do Conhecimento (SOCs), estudando os

vocabulários criados para controle de vocabulário e os vocabulários utilizados para estruturar os termos e permitir a descrição de conteúdos e de informações descritivas de recursos informacionais e de conjuntos de dados disponibilizados na Web de maneira formal, legíveis por máquina (Jesus; Santarém Segundo, 2023). Os autores também apontam a necessidade de formalização do processo de descrição de recursos informacionais e dos conjuntos de dados e metadados contidos na Web. O vocabulário em discussão neste tópico apresenta proposta para essa questão.

O *Data Catalog Vocabulary* (DCAT-3) figura como um vocabulário voltado à descrição de catálogos e seus conjuntos de dados, buscando favorecer a interoperabilidade na Web (Tomoyose, 2021). Segundo Lóscio, Burle e Calegari (2019), a utilização de vocabulários compartilhados para metadados (em especial os metadados estruturais, de procedência, de qualidade e de versionamento) auxilia o processo de comparação e o processamento automático dos dados e dos metadados.

O DCAT-3 é um vocabulário para a publicação de catálogos de dados na Web que foi originalmente desenvolvido no contexto de catálogos de dados governamentais, mas também é aplicável e tem sido usado em outros contextos (Albertoni *et al*, 2024, *online*). Na pesquisa de Albertoni *et al.* (2024), os autores conceituam e apresentam as versões desenvolvidas, conforme destacado nos parágrafos a seguir.

O vocabulário DCAT foi, inicialmente, desenvolvido e hospedado no *Digital Enterprise Research Institute* (DERI), depois, refinado pelo *eGov Interest Group* e, finalmente, padronizado em 2014 (VOCAB-DCAT-1) pelo *Government Linked Data (GLD) Working Group*.

Uma segunda revisão recomendada do DCAT, o DCAT-2 (VOCAB-DCAT-2), foi desenvolvida pelo *Dataset Exchange Working Group* em resposta a um novo conjunto de Casos de Uso e Requisitos (DCAT-UCR) coletados a partir da experiência das pessoas com o vocabulário DCAT desde a versão original, além de novas aplicações, que não foram consideradas na primeira versão.

O DCAT-3 foi desenvolvido pelo *Dataset Exchange Working Group*, considerando alguns dos casos de uso e algumas solicitações mais urgentes entre aqueles que não foram abordados na rodada anterior de padronização, e foi lançado como recomendação do W3C, em junho de 2024.

O DCAT-3 substitui seu antecessor, DCAT-2, porém não o torna obsoleto, uma vez que seus termos preservam a compatibilidade entre as versões. Os editores sugerem que as novas

implementações utilizem a nova versão, no entanto, as já existentes não precisam realizar atualizações, a não ser que queiram utilizar os novos recursos. Segundo eles, com o DCAT é possível:

> [...] que um publicador descreva conjuntos de dados e serviços de dados em um catálogo usando um modelo e vocabulário padrão que facilita o consumo e a agregação de metadados de múltiplos catálogos. Isso pode aumentar a descoberta de conjuntos de dados e serviços de dados. Também possibilita uma abordagem descentralizada para a publicação de catálogos de dados e torna possível a busca federada por conjuntos de dados em catálogos em múltiplos sites, utilizando o mesmo mecanismo e estrutura de consulta. Metadados agregados do DCAT podem servir como um arquivo de manifesto como parte do processo de preservação digital (Albertoni *et al*, 2024, *online*).

O DCAT-3 fornece classes e propriedades RDF para permitir que conjuntos de dados e serviços de dados sejam descritos e incluídos em um catálogo. O uso de um modelo e um vocabulário padrão facilita o consumo e a agregação de metadados de múltiplos catálogos, o que pode:

- Aumentar a recuperação de conjuntos de dados e serviços de dados;
- Permitir a busca federada por conjuntos de dados em catálogos em múltiplos *sites*;

O DCAT-3 utiliza-se de termos de outros vocabulários, como *Friend Of a Friend* (Foaf), *Ontology Web Language* (OWL), *Provenance Ontology* (Prov-O), *Vocabulary of Interlinked Datasets* (VoID) e, especialmente, do Dublin Core, sendo que o DCAT define um conjunto mínimo de propriedades e classes próprias (Albertoni *et al*, 2024, *online*). Os *namespaces* e prefixos usados nas partes normativas são apresentados no Quadro 6:

Quadro 6 - *Namespaces* do DCAT-3

Prefixo	Namespace IRI	Link de acesso
admns	http://www.w3.org/ns/adms#	https://www.w3.org/TR/vocab-adms/
dc	http://purl.org/dc/elements/1.1/	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
dcat	http://www.w3.org/ns/dcat#	https://www.w3.org/TR/vocab-dcat/
dcterms	http://purl.org/dc/terms/	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

dctype	http://purl.org/dc/dcmitype/	https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
foaf	http://xmlns.com/foaf/0.1/	http://xmlns.com/foaf/spec/
locn	http://www.w3.org/ns/locn#	http://www.w3.org/ns/locn
odrl	http://www.w3.org/ns/odrl/2/	https://www.w3.org/TR/odrl-vocab/
owl	http://www.w3.org/2002/07/owl#	https://www.w3.org/TR/owl2-syntax/
prov	http://www.w3.org/ns/prov#	https://www.w3.org/TR/prov-o/
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	https://www.w3.org/TR/rdf-syntax-grammar/
rdfs	http://www.w3.org/2000/01/rdf-schema#	https://www.w3.org/TR/rdf-schema/
skos	http://www.w3.org/2004/02/skos/core#	https://www.w3.org/TR/skos-reference/
spdx	http://spdx.org/rdf/terms#	https://spdx.org/rdf/spdx-terms-v2.2/
time	http://www.w3.org/2006/time#	https://www.w3.org/TR/owl-time/
vcard	http://www.w3.org/2006/vcard/ns#	https://www.w3.org/TR/vcard-rdf/
xsd	http://www.w3.org/2001/XMLSchema#	https://www.w3.org/TR/xmlschema11-2/

Fonte: Albertoni *et al* (2024)

Há, também, *namespaces* e prefixos não normativos mostrados no Quadro 7:

Quadro 7 - *Namespaces* não normativos do DCAT-3

Prefixo	Namespace IRI	Link de acesso
dqv	http://www.w3.org/ns/dqv#	https://www.w3.org/TR/vocab-dqv/
earl	http://www.w3.org/ns/earl#	https://www.w3.org/TR/EARL10-Schema/
geosparql	http://www.opengis.net/ont/geosparql#	https://docs.ogc.org/is/22-047r1/22-047r1.html
oa	http://www.w3.org/ns/oa#	https://www.w3.org/TR/annotation-vocab/
pav	http://purl.org/pav/	https://pav-ontology.github.io/pav/
sdmx-attribute	http://purl.org/linked-data/sdmx/2009/attribute#	https://www.w3.org/TR/vocab-data-cube/
sdo	https://schema.org/	https://schema.org/

xhv	http://www.w3.org/1999/xhtml/vocab#	https://www.w3.org/1999/xhtml/vocab

Fonte: Albertoni *et al* (2024)

Os vocabulários FOAF, VCard e Adms são apresentados brevemente abaixo, pois, como exposto acima, para a utilização plena do DCAT, é necessária a interação com outros vocabulários:

- *Friend Of Friends* (FOAF):

  A linguagem FOAF é definida como um dicionário de propriedades e classes nomeadas, utilizando a tecnologia RDF, do W3C. Ela foi criada para descrever pessoas, grupos e relacionamentos na Web. A organização é feita em classes e propriedades (Brickley; Miller, 2014, *online*).

  O objetivo principal é permitir que informações sobre pessoas e seus relacionamentos estejam interligadas na Web de forma descentralizada e em formatos que possam ser facilmente combinados. Por exemplo, um documento FOAF pode conter dados sobre quem são seus amigos, onde trabalha e interesses pessoais, e esses dados podem ser lidos e conectados com dados de outras fontes FOAF, formando uma rede.

  Resumidamente, FOAF serve como uma base para construir redes de informações sobre pessoas e suas conexões, permitindo que diferentes fontes contribuam com partes dessa rede de maneira padronizada e interoperável (Brickley; Miller, 2014, *online*).

- *vCard Ontology* (vCard):

  vCard é uma especificação criada para a descrição de pessoas e organizações, elaborada originalmente para o compartilhamento de cartões de visita digitais, permitindo que as informações sejam lidas e interpretadas por diversos *softwares* e dispositivos, como aplicativos de e-mail, *smartphones* e sistemas de gerenciamento de contatos (Iannella *et al,* 2014, *online).*

- *Asset Description Metadata Schema* (Adms)

  Adms é um perfil do DCAT usado para descrever ativos semânticos (ou, simplesmente, "ativos"), definidos como metadados altamente reutilizáveis e dados de referência (por exemplo, listas de códigos, taxonomias, dicionários, vocabulários) que são

> utilizados para o desenvolvimento de sistemas de governo eletrônico (Dekkers, 2013, *online*, tradução nossa). O Adms auxilia na descrição de documentos que servem como referência e ajuda a padronizar o compartilhamento e o uso de informações.
>
> Ao contrário de catálogos de dados comuns, com foco em conjuntos de dados que precisam ser processados, o Adms facilita o acesso a documentos e estruturas de referência que podem ser abertos e lidos diretamente. O DCAT foi projetado para facilitar a interoperabilidade entre catálogos de dados, estando o catálogo em si no centro do vocabulário. O Adms, por sua vez, é focado nos ativos que estão dentro de um catálogo (Dekkers, 2013, *online*, tradução nossa).

Além destes, o DCAT é compatível com diversos outros vocabulários que utilizam o RDF e podem ser combinados de acordo com as necessidades específicas de uma organização. Para estar em conformidade com os princípios do DCAT, um catálogo de dados precisa preencher os seguintes requisitos:

- O acesso aos dados deverá estar organizado em conjuntos de dados, distribuições, serviços de dados e séries de conjuntos de dados.
- Uma descrição em RDF do próprio catálogo, dos recursos catalogados correspondentes e das distribuições deverá estar disponível (mas a escolha da sintaxe RDF, protocolo de acesso e política de acesso não são exigidos).
- O conteúdo de todos os campos de metadados que estão no catálogo e que contêm dados sobre o próprio catálogo, os recursos catalogados correspondentes e as distribuições deverão estar incluídos nessa descrição em RDF e serem expressos usando as classes e propriedades apropriadas do DCAT, exceto onde não exista tal classe ou propriedade.
- Todas as classes e propriedades definidas no DCAT deverão ser usadas de uma maneira consistente com a semântica declarada.

O DCAT-3 é baseado em 7 classes principais (*Catalog, Cataloged Resource, Catalog Record, Dataset, Dataset Series, Distribution, Data Service*) e mais 8 classes adicionais (*Concept Scheme, Concept, Organization/Person, Relationship, Role, Period Of Time, Location, Checksum*), descritas no Quadro 8:

Quadro 8 - Classes do padrão DCAT-3

Nº	Superclasse	Classe	Descrição	Propriedades	
1	dcat:Dataset	dcat:Catalog	Representa um catálogo, que é um conjunto de dados em que cada item individual é um registro de metadados descrevendo algum recurso.	*catalog record*	*catalog*
				resource	*homepage*
				dataset	*themes*
				service	
2	-	dcat:Resource[1]	Representa um conjunto de dados, um serviço de dados ou qualquer outro recurso que possa ser descrito por um registro de metadados em um catálogo. Não é para ser usada diretamente, é a superclasse que abrange dcat:Dataset, dcat:DataService e dcat:Catalog.	*access rights*	*landing page*
				conforms to	*qualified attribution*
				contact point	*license*
				creator	*rights*
				description	*has part*
				title	*has policy*
				release date	*is reference by*
				update / modification date	*previous version*
				language	*has version*
				publisher	*current version*
				identifier	*replaces*
				theme / category	
				type / genre	*version*

[1] Os recursos em um catálogo devem ser instâncias de uma dessas classes: dcat:Dataset, dcat:DataService, dcat:Catalog, ou de uma subclasse dessas, ou de uma subclasse de dcat:Resource definida em um perfil DCAT ou outra aplicação DCAT.

				relation	*version notes*
				qualified relation	*first*
				keyword / tag	*last*
				status	*previous*
3	dcat:Resource[1]	dcat:Dataset	Representa uma coleção de dados, publicada ou curada por um único agente ou comunidade identificável.	*distribution*	*spatial resolution*
				frequency	*temporal coverage*
				in series	*temporal resolution*
				spatial / geographical coverage	*was generated by*
4	-	dcat:Distribution	Representa uma forma acessível de um conjunto de dados, como um arquivo para *download*.	*title*	*download URL*
				description	*byte size*
				release date	*spatial resolution*
				update / modification date	*temporal resolution*
				license	*conforms to*
				access rights	*media type*
				rights	*format*
				has policy	*compression format*
				access URL	*packaging format*
				access service	*checksum*

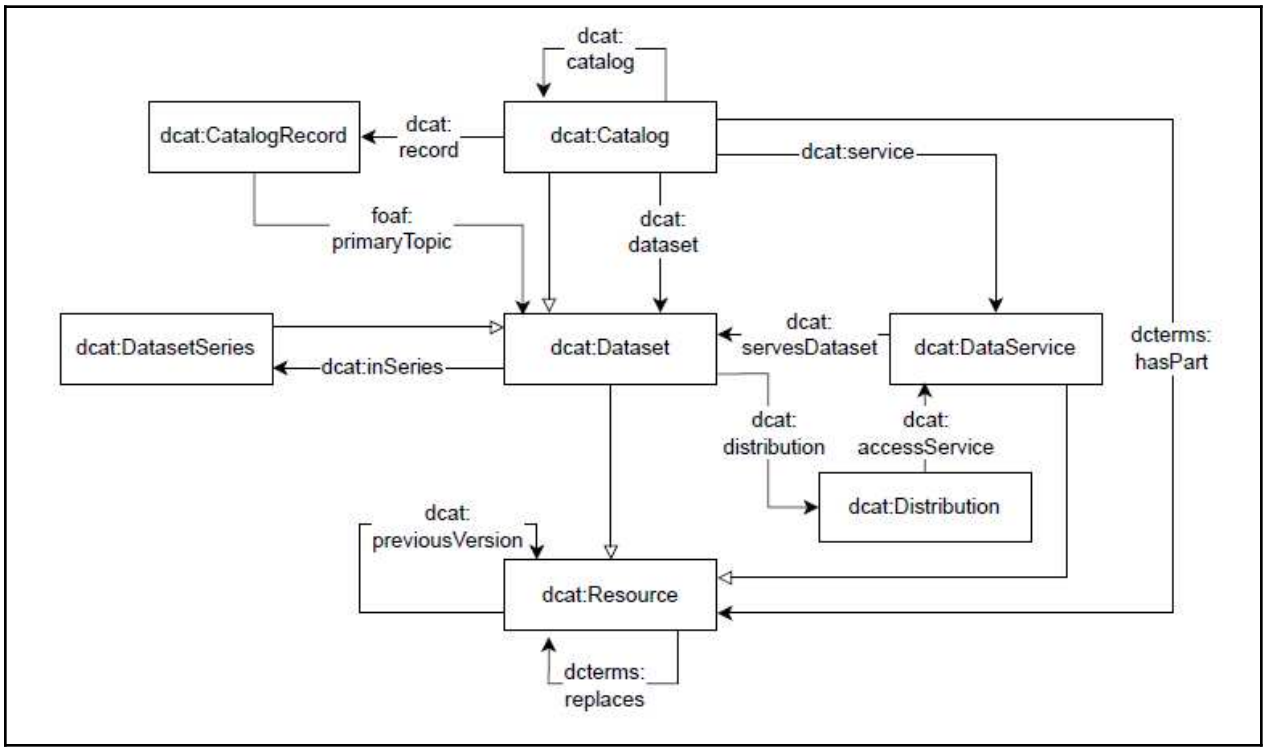
5	dcat:Resource dctype:Service	dcat:DataService	Representa uma coleção de operações acessíveis por meio de uma interface (API - *Application Programming Interface*) que fornecem acesso a um ou mais conjuntos de dados ou funções de processamento de dados.	*endpoint URL*	*serves dataset*
				endpoint description	
6	dcat:Dataset	dcat:DatasetSeries	Representa um conjunto de dados que representa uma coleção de conjuntos de dados publicados separadamente, mas que compartilham algumas características que os agrupam.		
7	-	dcat:CatalogRecord	Representa um registro de metadados no catálogo, principalmente relacionado às informações de registro, como quem adicionou o registro e quando.	*title*	*update / modification date*
				description	*primary topic*
				listing date	*conforms to*
8	-	skos:ConceptScheme	Um sistema de organização do conhecimento (SKOS) usado para representar temas/categorias de conjuntos de dados no catálogo.	-	
9	-	skos:Concept	Uma categoria ou um tema usado para descrever conjuntos de dados no catálogo.	-	
10	foaf:Agent	foaf:Person foaf:Organization	O vocabulário FOAF possui diversas propriedades para descrever essas entidades.		
11	prov:EntityInfluence	dcat:Relationship	Uma classe associada para inserir informações adicionais a um relacionamento entre recursos DCAT.	*relation*	*had role*
12	skos:Concept	dcat:Role	Usado para especificar o papel de um Agente em relação a uma Entidade.	-	
13	-	dcterms:PeriodofTime	Um intervalo de tempo que é nomeado ou definido por seu início e fim.	*start date*	*beginning*
				end date	*end*
14	-	dcterms:Location	Uma região espacial ou localidade.	*geometry*	*centroid*

				bounding box	
15	-	spdx:Checksum	Permite que os resultados de uma variedade de algoritmos de *checksum* e de resumo criptográfico de mensagens sejam representados.	*algorithm*	*checksum value*

Fonte: adaptado de Albertoni *et al* (2024)

Geißner (2023), apresenta uma imagem simplificada das relações entre as principais classes do DCAT-3, reproduzida na Figura 1:

Figura 1 - Relações entre as sete classes principais do DCAT-3



Fonte: Geißner (2023, p. 17)

O DCAT-3 foi pensado para o contexto dos catálogos e seus dados e, para isso, apresenta uma estrutura básica para a descrição dos componentes de um catálogo, para a representação dos conjuntos de dados, para a representação de cada versão de um recurso, entre outras possibilidades a serem avaliadas de acordo com cada contexto específico. Na próxima seção, será abordada a Governança de Dados, que guia todo o processo para a gestão no ciclo de vida dos dados.

3.4 Governança de Dados

Na era dos dados, esses elementos precisam ser gerenciados como ativos organizacionais (Barbieri, 2019), e um conceito que guia a questão administrativa do tratamento de dados é o da

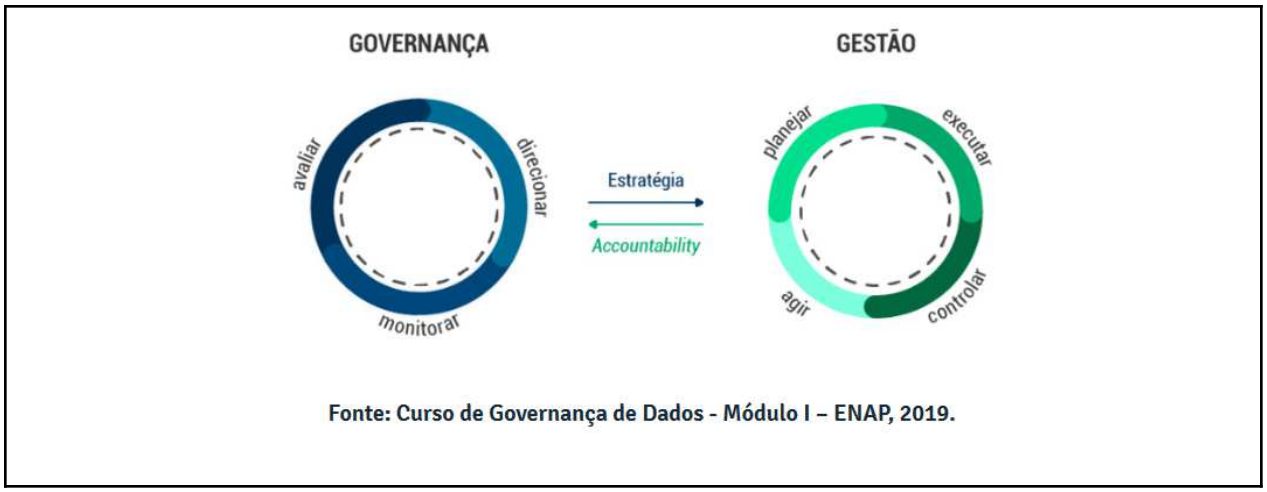
governança de dados, que pode ser definida como uma orientação sobre ações de gerenciamento de dados nas perspectivas de compartilhamento, arquitetura, segurança, qualidade, operação, entre outros aspectos tecnológicos (Brasil, s.d.g, *online*). Essas ações podem ser implementadas por meio de políticas, acordos, papéis e definição de métodos a serem utilizados nas atividades de criação, armazenamento, avaliação, uso e eliminação de dados (Espíndola *et al*, 2018*)*, e uma das formas mais comuns para adotar a governança de dados é a partir de projetos especiais de dados (Barbieri, 2019).

Segundo Barbieri (2019), esse termo surgiu a partir do conceito geral de 'governança' e foca nos princípios de organização e controle de dados para a produção de informação e conhecimento das empresas. A governança de dados também abrange, além dos dados, questões relacionadas à melhor organização e integração de metadados das organizações, transformando-os em aderentes às necessidades do negócio, legíveis e disponíveis para acesso (Fraga, 2022). É um conceito que estabelece as diretrizes e as políticas a serem seguidas pela gestão de dados, que se refere a processos organizacionais, responsabilizando-se pelo planejamento, pela execução, pelo controle e pelo monitoramento (Brasil, 2022).

Para estabelecer as primeiras camadas de governança de dados, é necessário definir uma estrutura corporativa formal, composta por elementos de negócios e de TI, regida por políticas amplas de dados (Barbieri, 2019). Além disso, segundo o autor, é fundamental definir papéis de gestores de dados (*data stewards*) e de responsáveis por dados (*data owners*), que serão referências responsáveis (*accountability*) por um ativo específico dentro do contexto organizacional.

A governança de dados, ao acompanhar a gestão de dados na organização, garante que os dados gerados estejam alinhados aos objetivos organizacionais, além de monitorar e acompanhar os dados para possibilitar sua utilização de forma eficaz e eficiente, atendendo aos objetivos organizacionais traçados pela alta administração (Espíndola *et al,* 2018). A Figura 2 demonstra as particularidades de cada processo (governança e gestão) de forma integrada, em que, no ciclo da governança, há ações de direcionamento, monitoramento e avaliação, e na gestão, o ciclo *Plan, Do, Check, Act* (PDCA):

Figura 2 - Relação entre governança de dados e gestão de dados



Fonte: Brasil (2022).

Alguns riscos da não implementação de uma política de governança de dados (um conjunto de estratégias para implementação da governança de dados) em uma organização são: risco de não conformidade em relação a regulações de privacidade e de sigilo, risco de indisponibilidade de dados e risco de baixa qualidade de dados (Brasil, 2022). A governança de dados auxilia no fomento, no aprimoramento de políticas públicas e no fornecimento de soluções e serviços ao cidadão (Brasil, s.d.g, *online*) e, por isso, pode ser considerada fundamental no contexto dos DGA.

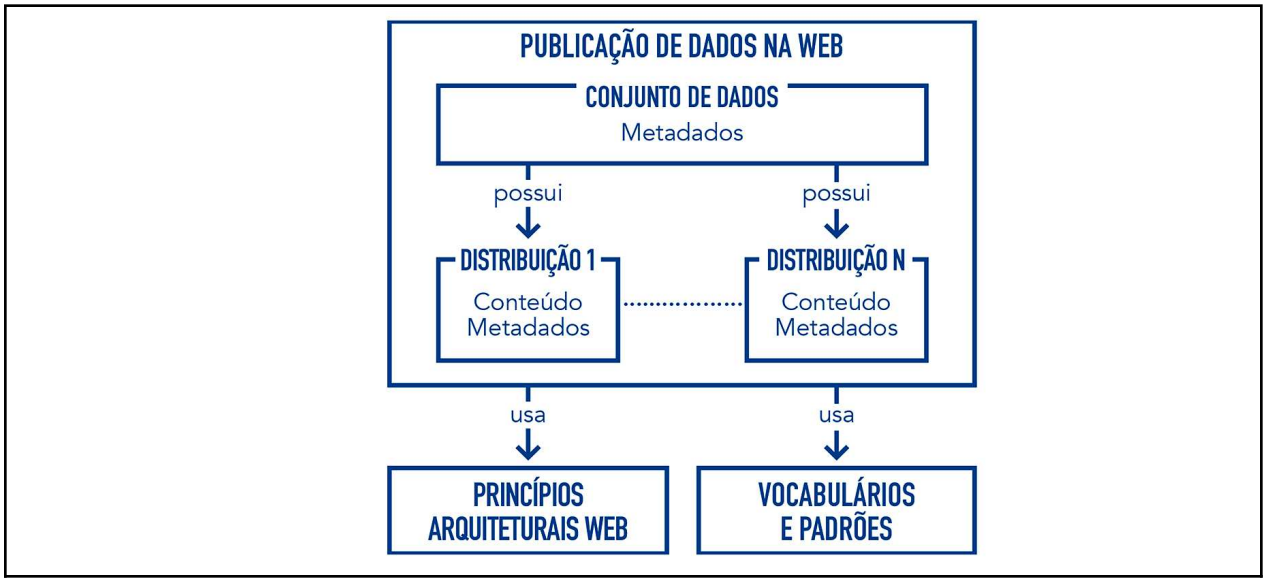
Sendo um dos princípios da governança o estabelecimento de diretrizes para gestão dos dados, apresentam-se, na próxima subseção, as Boas Práticas para Dados na Web, que são um conjunto de recomendações para publicação de dados na Web, aplicadas, neste trabalho, aos dados abertos e utilizadas para embasamento da proposta de metadados, apresentada na Seção 6.

3.5 Boas Práticas para Dados na Web (*Data on the Web Best Practices*)

As Boas Práticas Para Dados na Web são uma recomendação do W3C, de 31 de janeiro de 2017, com um conjunto de 35 boas práticas com o objetivo de estimular e possibilitar a expansão da Web como um ambiente para a troca de dados e possuem uma tradução autorizada de 30 de outubro de 2019, que foi utilizada para apresentar a fundamentação desta seção (Lóscio; Burle; Calegari, 2019).

Um elemento fundamental na publicação e no compartilhamento de dados na Web é a sua base arquitetônica. O princípio da identificação estabelece que *Uniform Resource Identifier* (URI) devem ser utilizados para identificar recursos, o que se alinha a essa arquitetura. Neste contexto, recursos podem representar tanto um conjunto de dados completo quanto um item específico de um conjunto. Todos os recursos precisam ser publicados com URIs estáveis, possibilitando sua referência e a criação de conexões entre múltiplos recursos por meio dos URIs. Além disso, para fomentar a interoperabilidade entre conjuntos de dados, é essencial adotar vocabulários padronizados e normas reconhecidas. A Figura 3 apresenta a estrutura dessa arquitetura:

Figura 3 - Arquitetura para publicação de dados na web



Fonte: Lóscio *et al* (2018).

Em duas revisões de literatura realizadas (Oliveira; Nascimento Silva, 2024 e Etapa 1 desta dissertação), evidenciaram-se diversos parâmetros e métricas que podem ser aplicados a dados abertos, como as 5 estrelas de Tim Berners Lee, a métrica DGA Br e os princípios FAIR, e selecionaram-se as Boas Práticas para Dados na Web pois elas apresentam diversos aspectos que se relacionam com os metadados, objeto central dessa proposta.

Um resumo das 35 Boas Práticas para Dados na Web é apresentado no Quadro 9 e destacado em negrito com fundo azul claro, constam as boas práticas relacionadas à metadados:

Quadro 9 - Boas Práticas para Dados na Web

Categoria	ID	Descrição	Benefícios
Metadados	**Boa prática 1 - Fornecer metadados**	**Fornecer metadados tanto para usuários pessoas quanto para aplicações de computadores**	**Reúso, Compreensão, Facilidade de descoberta, Facilidade de processamento**
	Boa prática 2 - Fornecer metadados descritivos	**Fornecer metadados que descrevam as características gerais dos conjuntos de dados e das distribuições**	**Reúso, Compreensão, Facilidade de descoberta**
	Boa prática 3 - Fornecer metadados estruturais	**Fornecer metadados que descrevam o esquema e a estrutura interna de uma distribuição**	**Reúso, Compreensão, Facilidade de processamento**
Licença de Dados	**Boa prática 4 - Fornecer informações sobre a licença de dados**	**Fornecer um *link* ou uma cópia dos termos da licença que controla a utilização dos dados**	**Reúso, Confiança**
Procedência de Dados	**Boa prática 5 - Fornecer informações de procedência de dados**	**Fornecer informações completas sobre as origens dos dados e de quaisquer alterações que você tenha feito.**	**Reúso, Compreensão, Confiança**
Qualidade de Dados	Boa Prática 6 - Fornecer informações de qualidade de dados	Fornecer informações sobre a qualidade dos dados e adequação para fins específicos	Reúso, Confiança
Versionamento de Dados	**Boa prática 7 - Fornecer indicador de versão**	**Designar e indicar um número de versão ou data para cada conjunto de dados**	**Reúso, Confiança**
	Boa prática 8 - Fornecer o histórico de versão	**Fornecer um histórico completo de versão que explique as alterações feitas em cada versão**	**Reúso, Confiança**
Identificadores de Dados	**Boa prática 9 - Usar URIs persistentes como identificadores de conjuntos de dados**	**Identificar cada conjunto de dados por meio de um URI persistente e cuidadosamente escolhido**	**Reúso, Capacidade de conexão, Interoperabilidade**
	Boa Prática 10: Usar URIs persistentes como identificadores dentro de conjuntos de dados	Reutilizar os URIs de terceiros como identificadores dentro dos conjuntos de dados quando possível	Reúso, Capacidade de conexão, Facilidade de descoberta, Interoperabilidade

	Boa Prática 11: Atribuir URIs para as versões dos conjuntos de dados e séries	Atribuir URIs a versões individuais de conjuntos de dados, assim como a séries em geral	Reúso, Facilidade de descoberta, Confiança
Formatos de Dados	**Boa prática 12 - Usar formatos de dados padronizados legíveis por máquinas**	**Disponibilize os dados em um formato padronizado e legível por máquinas que seja apropriado para seu propósito ou uso em potencial**	**Reúso, Facilidade de processamento**
	Boa prática 13 - Usar representações de dados que sejam independentes de localidade (*locale neutral*)	**Utilize estruturas e conteúdo de dados de localidade neutra ou, quando isso não for possível, forneça metadados sobre os parâmetros regionais dos conteúdos de dados**	**Reúso, Compreensão**
	Boa prática 14 - Fornecer dados em formatos múltiplos	**Disponibilize os dados em vários formatos quando mais de um formato for adequado ao potencial ou pretendido uso do dado**	**Reúso, Facilidade de processamento**
Vocabulários de dados	**Boa Prática 15 - Reutilizar vocabulários, dando preferência aos padronizados**	**Utilize termos oriundos de vocabulários compartilhados, preferencialmente os padronizados, para codificar dados e metadados**	**Reúso, Facilidade de processamento, Compreensão, Confiança, Interoperabilidade**
	Boa Prática 16: Escolher o nível de formalização adequado	Escolha um nível de semântica formal que se ajuste tanto aos dados quanto às aplicações mais prováveis de serem utilizadas	Reúso, Compreensão, Interoperabilidade
Acesso a dados	Boa Prática 17: Fornecer *download* em massa *(bulk download)*	Permitir que os consumidores acessem o conjunto de dados completo em uma única solicitação	Reúso, Facilidade de acesso
	Boa Prática 18: Fornecer subconjuntos para conjuntos de dados extensos	Caso seu conjunto de dados seja grande, permitir que usuários e aplicações trabalhem prontamente por meio do uso de subconjuntos de seus dados	Reúso, Capacidade de conexão, Facilidade de acesso, Facilidade de processamento
	Boa Prática 19: Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos	Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos, para além de extensões de arquivos	Reúso, Facilidade de acesso

	Boa Prática 20: Fornecer acesso em tempo real	**Quando os dados forem produzidos em tempo real, disponibilizá-los na Web em tempo real, ou quase em tempo real**	**Reúso, Facilidade de acesso**
	Boa Prática 21: Fornecer dados atualizados	**Disponibilizar dados de forma atualizada e explicitar a frequência de atualização**	**Reúso, Facilidade de acesso**
	Boa Prática 22: Fornecer uma explicação para os dados que não estão disponíveis	**Para dados que não estão disponibilizados, fornecer uma explicação sobre como os dados podem ser acessados e quem pode fazê-lo**	**Reúso, Confiança**
APIs de acesso a dados	**Boa Prática 23: Disponibilizar dados por meio de uma API**	**Disponibilizar uma API para servir os dados, caso você tenha recursos para tanto**	**Reúso, Facilidade de processamento, Interoperabilidade, Facilidade de acesso**
	Boa Prática 24: Usar padrões Web como base para construção de APIs	Ao desenvolver APIs, utilizar um estilo arquitetônico baseado nas tecnologias da própria Web	Reúso, Capacidade de conexão, Facilidade de descoberta, Facilidade de acesso, Facilidade de processamento
	Boa Prática 25: Fornecer documentação completa para as APIs	Fornecer informações completas na Web sobre a API. Atualizar a documentação conforme características adicionadas ou modificações realizadas	Reúso, Confiança
	Boa Prática 26: Evitar alterações que afetem o funcionamento de sua API	Evitar alterações em sua API que afetem o código do cliente e, quando houver evolução, informar seus desenvolvedores sobre quaisquer modificações feitas na API	Confiança, Interoperabilidade
Preservação de dados	Boa Prática 27: Preservar identificadores	Ao remover dados da Web, preservar o identificador e fornecer informações sobre o recurso arquivado	Reúso, Confiança
	Boa Prática 28: Avaliar a cobertura do conjunto de dados	Avaliar a cobertura de um conjunto de dados antes de sua preservação	Reúso, Confiança

Feedback	**Boa Prática 29: Coletar *feedback* de consumidores de dados**	**Fornecer meios fáceis de se encontrar para que os consumidores deem *feedback***	**Reúso, Compreensão, Confiança**
	Boa Prática 30: Compartilhar o *feedback* disponível	Disponibilizar publicamente mecanismos de *feedback* de consumidor sobre conjuntos de dados e distribuições	Reúso, Confiança
Enriquecimento de dados	**Boa Prática 31: Enriquecer dados por meio da geração de novos dados**	**Enriqueça seus dados gerando novos dados, pois, ao fazê-lo, você estará aumentando o seu valor**	**Reúso, Compreensão, Confiança, Facilidade de processamento**
	Boa Prática 32: Fornecer visualizações complementares	Enriqueça seus dados apresentando-os em formas complementares e diretamente informativas, tais como visualizações, tabelas, aplicações Web ou resumos	Reúso, Compreensão, Facilidade de acesso, Confiança
	Boa Prática 33: Fornecer *feedback* para o publicador original	Informar ao publicador original quando você está promovendo o reúso de seus dados. Informe-o caso encontre um erro ou tenha sugestões ou elogios a tecer	Reúso, Interoperabilidade, Confiança
	Boa Prática 34: Obedecer os termos de licença	Encontrar e seguir os requisitos de licença indicada pelo publicador original do conjunto de dados	Reúso, Confiança
Republicação de dados	**Boa Prática 35: Citar a publicação original do conjunto de dados**	**Reconhecer a fonte de seus dados nos metadados. Caso forneça uma interface de usuário, inclua a citação claramente na interface**	**Reúso, Facilidade de descoberta, Confiança**

Fonte: adaptado de Lóscio, Burle e Calegari (2019)

Cada benefício listado na última coluna do Quadro 9 representa uma melhoria na maneira como os conjuntos de dados são disponibilizados na Web e, abaixo, encontra-se a descrição de cada um:

- **Compreensão**: as pessoas terão um melhor entendimento sobre a estrutura e o significado dos dados, os metadados e a natureza do conjunto de dados.

- **Facilidade de Processamento**: as máquinas serão capazes de processar e manipular automaticamente os dados dentro de um conjunto de dados.
- **Facilidade de Descoberta**: as máquinas serão capazes de descobrir automaticamente um conjunto de dados ou dados dentro de um conjunto de dados.
- **Reúso**: aumentarão as possibilidades de reutilização do conjunto de dados por diferentes grupos de consumidores de dados.
- **Confiança**: a confiança que os clientes depositam no conjunto de dados se ampliará.
- **Capacidade de Conexão**: será possível criar conexões entre os recursos de dados (conjuntos de dados e itens de dados).
- **Facilidade de Acesso**: pessoas e máquinas serão capazes de acessar dados atualizados de várias maneiras.
- **Interoperabilidade**: será mais fácil alcançar um consenso entre publicadores de dados e consumidores (Lóscio; Burle; Calegari, 2019).

Ao todo, foram identificadas 19 Boas Práticas que podem ser relacionadas com campos de metadados, representando aproximadamente 50% do total, e são utilizadas como embasamento na discussão da seção 6. Na próxima seção, serão apresentados os procedimentos realizados para a concretização da proposta desta dissertação.

4 PROCEDIMENTOS METODOLÓGICOS

Esta seção apresenta a caracterização da pesquisa e o detalhamento das etapas que compõem o estudo. A pesquisa classifica-se como aplicada, adotando uma abordagem mista, adequada para investigar a representação dos DGA, e segue um delineamento exploratório e descritivo, utilizando-se principalmente de técnicas de pesquisa bibliográfica e observação direta para atingir os objetivos propostos.

4.1 Caracterização da pesquisa

Delineiam-se, a seguir, a caracterização e o percurso metodológico adotados para conduzir a presente pesquisa, explicitando os procedimentos, métodos e técnicas empregados para alcançar os objetivos deste projeto de pesquisa. Lakatos e Marconi (2003) definem pesquisa como um procedimento formal, que requer tratamento científico, constituindo meios para o conhecimento da realidade ou para desvendar verdades parciais.

A pesquisa em questão caracteriza-se como **aplicada**, com a finalidade de resolver problemas reais; **exploratória**, objetivando desenvolver e esclarecer conceitos e proporcionar uma visão geral sobre um fato; e **descritiva**, visando descrever características de determinado fenômeno e estabelecer relações entre variáveis (Gil, 2017). A pesquisa de natureza aplicada é aquela que se debruça sobre problemas que necessitam de soluções práticas. Uma de suas características fundamentais é o interesse na aplicação e utilização dos conhecimentos (Gil, 2008), envolvendo verdades e interesses regionais (Prodanov; Freitas, 2013).

Em relação aos objetivos, a pesquisa exploratória tem o propósito de propiciar maior familiaridade com um problema, buscando torná-lo mais evidente ou construir hipóteses (Gil, 2017). Considera-se que é uma pesquisa em fase preliminar, envolvendo levantamento bibliográfico, entrevistas com pessoas que tiveram experiências práticas com o problema e análise de exemplos que estimulem a compreensão do fato (Prodanov; Freitas, 2013). Para Severino (2013), esse tipo de pesquisa dispõe-se a levantar informações sobre determinado objeto e delimitar um campo de trabalho. Em uma explanação completa, Gil (2008) apresenta a definição:

> As pesquisas exploratórias têm como principal finalidade desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores. [...]
> Pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato. Este tipo de pesquisa é realizada especialmente quando o tema escolhido é pouco explorado e torna-se difícil sobre ele formular hipóteses precisas e operacionalizáveis.
> Muitas vezes as pesquisas exploratórias constituem a primeira etapa de uma investigação mais ampla. Quando o tema escolhido é bastante genérico, tornam-se necessários seu esclarecimento e delimitação, o que exige revisão da literatura, discussão com especialistas e outros procedimentos. O produto final deste processo passa a ser um problema mais esclarecido, passível de investigação mediante procedimentos mais sistematizados (Gil, 2008, p. 27).

Ainda em relação aos objetivos, essa pesquisa identifica-se também como descritiva, sendo aquela em que o pesquisador apenas registra e descreve fatos observados, sem interferência, e envolve técnicas padronizadas de coleta de dados. Juntamente com as pesquisas exploratórias, são aquelas habitualmente realizadas por pesquisadores sociais envolvidos com a atuação prática (Gil, 2008; Prodanov; Freitas, 2013). Complementa-se essa definição na visão de Prodanov e Freitas:

> A pesquisa descritiva procura descobrir a frequência com que um fato ocorre, sua natureza, suas características, causas, relações com outros fatos.
> Assim, para coletar tais dados, utiliza-se de técnicas específicas, dentre as quais se destacam a entrevista, o formulário, o questionário, o teste e a observação.
> A pesquisa descritiva procura classificar, explicar e interpretar fatos que ocorrem.
> Nas pesquisas descritivas, os fatos são observados, registrados, analisados, classificados e interpretados, sem que o pesquisador interfira sobre eles, ou seja, os fenômenos do mundo físico e humano são estudados, mas não são manipulados pelo pesquisador (Prodanov; Freitas, 2013, p. 52).

O objeto de estudo desta dissertação aborda uma temática recente e ainda pouco explorada, assim, um dos procedimentos utilizados, a fim de analisar o estado da arte e situar o problema de pesquisa, foi a realização de revisões de literatura, consideradas por Prodanov e Freitas (2013) como fundamentais no trabalho acadêmico, sendo a partir destas que se contextualiza o trabalho dentro da grande área de pesquisa da qual faz parte. A revisão caracteriza-se por utilizar dados que já foram estudados por outros pesquisadores e estão registrados, trabalhando a partir das contribuições dos autores dos estudos analíticos constantes nas bibliografias (Severino, 2013).

Segundo Lakatos e Marconi (2003), qualquer pesquisa científica se inicia com o levantamento de dados, realizado de duas maneiras: a pesquisa documental (fontes primárias), e

pesquisa bibliográfica (fontes secundárias). As autoras caracterizam a pesquisa documental como aquela em que a fonte de coleta de dados é restrita a documentos, escritos ou não, e a qual pode ser feita no momento em que o fato ou fenômeno ocorre, ou depois. Quanto à pesquisa bibliográfica, elas a consideram como aquela que:

> Abrange toda bibliografia já tornada pública em relação ao tema de estudo, desde publicações avulsas, boletins, jornais, revistas, livros, pesquisas, monografias, teses, material cartográfico etc., até meios de comunicação orais: rádio, gravações em fita magnética e audiovisuais: filmes e televisão. Sua finalidade é colocar o pesquisador em contato direto com tudo o que foi escrito, dito ou filmado sobre determinado assunto, inclusive conferências seguidas de debates que tenham sido transcritos por alguma forma, quer publicadas, quer gravadas (Lakatos; Marconi, 2003, p. 183).

Na primeira parte do presente trabalho, foi utilizada principalmente a pesquisa bibliográfica, sendo elaborada a partir de material já publicado, podendo incluir “material impresso, como livros, revistas, jornais, teses, dissertações e anais de eventos científicos" (Gil, 2017, p. 33).

Para atingir o primeiro objetivo específico, foi realizada uma revisão de literatura, em que foi definido um determinado problema e realizada a sumarização de estudos prévios, apresentando aos leitores o panorama da área de investigação. A revisão de literatura tem o objetivo de identificar lacunas e apresentar sugestões para a resolução de problemas (Koller, Couto, Von Hohendorff, 2014). Segundo Gil (2017), considera-se esse levantamento bibliográfico preliminar como um estudo exploratório, com a finalidade de familiarizar o pesquisador com a área de estudo de interesse e sua delimitação, sendo essencial para a formulação clara e precisa do problema de pesquisa.

Por fim, caracteriza-se essa pesquisa de acordo com sua abordagem, do tipo **qualitativa e quantitativa**. A abordagem qualitativa considera o processo de análise dos dados e seus significados. A interpretação dos fenômenos e a atribuição de significados são focos principais de abordagem, não tendo a prioridade de numerar ou medir unidades (Severino, 2013). Esse autor ainda explica que os dados coletados neste tipo de pesquisa são descritivos e retratam o maior número possível de elementos que estão presentes na realidade estudada. Em relação à abordagem quantitativa, Prodanov e Freitas (2013) afirmam que é possível traduzir em número opiniões e informações para classificá-las e analisá-las, e que é necessária a utilização de recursos e de técnicas estatísticas.

A pesquisa foi realizada em quatro etapas principais, correspondendo aos quatro objetivos específicos detalhados na subseção 1.2.2, conforme apresentado no Quadro 10, com o detalhamento de etapas, técnicas, procedimentos e artefatos:

Quadro 10 - Delineamento Metodológico da Pesquisa

ETAPA	OBJETIVO ESPECÍFICO	NATUREZA	ABORDAGEM METODOLÓGICA	FINALIDADE DA PESQUISA	TÉCNICAS E INSTRUMENTOS	PROCEDIMENTOS	ARTEFATO
Etapa 1: revisão integrativa de literatura	1- Identificar, na literatura acadêmica brasileira, modelos, métricas, *frameworks* e instrumentos similares que abordem a publicação, a avaliação, o ciclo de vida ou o tratamento de dados abertos ou DGA, a fim de evidenciar abordagens, técnicas e métodos utilizados	APLICADA	Qualitativa Quantitativa	Exploratória e Descritiva	Pesquisa bibliográfica Revisão integrativa de literatura	Construção do protocolo; Execução do protocolo; Análise de conteúdo dos resultados qualitativos e, brevemente, dos quantitativos; Mapeamento de trabalhos correlatos à proposta da dissertação.	Seção 5.1
Etapa 2: mapeamento de iniciativas de dados abertos do governo federal	2- Mapear documentos e legislações (leis, decretos, normativas, portarias, resoluções e instrumentos) sobre os dados abertos no governo federal brasileiro		Qualitativa	Descritiva	Pesquisa documental Pesquisa bibliográfica	Levantamento de legislações, normativas, portarias e instrumentos relacionados a dados abertos; Análise dos documentos selecionados sob a ótica da CI.	Seção 2 Seção 5.2.1
Etapa 3: mapeamento formatos e padrões de publicação no contexto brasileiro	3- Identificar e analisar as tipologias e os formatos de dados abertos publicados, os padrões e os metadados utilizados no Portal Brasileiro de Dados Abertos		Qualitativa	Exploratória e Descritiva	Pesquisa documental Pesquisa bibliográfica Observação direta	Selecionar os instrumentos da Política de Dados Abertos do Brasil; Analisar as tipologias de dados abertos, formatos, padrões e metadados utilizados; Identificar lacunas e pontos de melhoria.	Seção 2 Seções 5.2.2 e 5.2.3
Etapa 4: elaboração de proposta de metadados para representação de dados abertos do governo federal	4 - Propor uma estrutura de representação descritiva para os dados abertos do governo federal brasileiro no Portal Brasileiro de Dados Abertos com a utilização de metadados		Qualitativa	Descritiva	Observação direta e análise dos resultados	Sistematização e discussão dos resultados das Etapas 1, 2 e 3; Elaboração de proposta de representação contendo metodologia e recomendações para a representação de dados abertos do Governo Federal com a utilização de metadados.	Seção 6 - Proposta de recomendação e modelo de metadados

Fonte: elaborado pela autora (2024).

4.2 Etapas da pesquisa

A seguir, são detalhadas as etapas que constituem o desenvolvimento desta pesquisa, organizadas de maneira sequencial, de forma a apresentar uma visão estruturada do processo adotado.

4.2.1 Etapa 1: revisão integrativa de literatura

O objeto de estudo desta dissertação integra uma temática recente e ainda pouco explorada e visando analisar o primeiro objetivo específico: "**Identificar, na literatura acadêmica brasileira, modelos, métricas, *frameworks* e instrumentos similares que abordem a publicação, a avaliação, o ciclo de vida ou o tratamento de dados abertos ou DGA, a fim de evidenciar abordagens, técnicas e métodos utilizados**", foi criado um protocolo com os critérios utilizados para realizar uma revisão integrativa de literatura, que possui uma ampla abordagem metodológica, permitindo a inclusão de estudos experimentais e não-experimentais, de forma a permitir uma compreensão completa do fenômeno analisado e combinando dados da literatura teórica e empírica (Souza; Silva; Carvalho, 2010).

Antes de iniciar o detalhamento dos procedimentos utilizados para a revisão integrativa, é importante destacar que essa necessidade foi identificada após uma revisão de literatura realizada para conhecer abordagens sobre representação e recuperação dos DGA na literatura científica nacional e internacional em que Oliveira e Nascimento Silva (2024) identificaram que a representação dos DGA é um elemento-chave para garantir a qualidade desses dados e satisfazer as necessidades dos usuários de sistemas de informação. Além disso, nesse estudo, a representação foi relacionada com a publicação, a recuperação e o acesso, sendo todos eles essenciais para a reutilização desses dados.

Ainda nessa pesquisa, os vocabulários controlados são considerados uma boa prática para a representação e a disponibilização dos DGA, por serem um fator decisivo na interconexão e na interoperabilidade dos dados, e os metadados foram considerados essenciais para a representação em ambientes digitais, sendo também relacionados à qualidade. Com isso, uma boa definição desses metadados implicará diretamente na melhor caracterização dos DGA, abordagem realizada nessa dissertação.

Uma lacuna apontada por Oliveira e Nascimento Silva (2024) é a carência de abordagens visando ao usuário consumidor, especialmente ao não especialista. Esse hiato é abordado sob uma perspectiva no presente trabalho, ao pensar em uma representação que beneficie justamente esse tipo de usuário, porém, com a limitação de não investigar diretamente as necessidades desses usuários e também de não realizar validação com eles, o que representa uma oportunidade para trabalhos futuros.

As perspectivas do trabalho de Oliveira e Nascimento Silva (2024) foram ratificadas na revisão integrativa de literatura executada na Etapa 1 desta dissertação, em que as conclusões sobre os desafios representados pela quantidade e heterogeneidade dos dados disponíveis foram semelhantes. O presente trabalho também debruça-se sobre essa lacuna, assim como sobre a possível solução proposta, que é a descrição e a representação adequada para amenizar essas questões.

Protocolo da Revisão Integrativa

Inicialmente, foi elaborado um protocolo baseado em autores sobre a temática de revisão de literatura, como Arlene Fink (2019), Galvão e Pereira (2014) e Brito e Martins (2023), considerando as seguintes definições iniciais do protocolo: fontes de dados, estratégia de busca, critérios de elegibilidade, inclusão e exclusão e análise dos documentos.

Na seleção das **fontes de dados**, optou-se por utilizar a base de pesquisa que integra os sistemas de informação de teses e dissertações existentes nas instituições de ensino e pesquisa do Brasil, que é a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). A BDTD é um portal de busca com o objetivo de prover acesso a textos completos de teses e dissertações defendidas em diversas instituições brasileiras de ensino e pesquisa, de forma gratuita. Essa fonte foi selecionada como um recorte para os trabalhos acadêmicos produzidos no cenário nacional, após investigação do cenário nacional e internacional no trabalho de Oliveira e Nascimento Silva (2024).

Para a construção da **estratégia de busca**, primeiramente foram realizados testes preliminares na BDTD, a fim de se conhecer os instrumentos de busca, e foram escolhidos os descritores ‘métrica’, ‘modelo’, ‘*framework*’, ‘dados governamentais abertos’ e suas variações. Após a realização dos testes, foi construída uma única *string* utilizando os operadores booleanos

AND e OR para fazer relacionamento entre os termos selecionados. A base permite pesquisar com as opções 'todos os termos', 'qualquer termo' ou 'nenhum termo'. A segunda opção foi a escolhida para ampliar a busca.

Depois da construção da *string,* foram estabelecidos os **critérios de elegibilidade, inclusão e exclusão** para os documentos recuperados. Foram filtrados os documentos em inglês, espanhol e português, sem delimitação temporal e, pela própria característica da base, apenas teses e dissertações. O resultado dessa etapa foi exportado e agrupado em *softwares* de planilhas eletrônicas, os resultados foram ordenados e comparados para a verificação de duplicatas e depois foram aplicados os critérios de inclusão e exclusão, sendo eles: idioma (português, inglês ou espanhol), delimitação de área de concentração (Ciência da Informação, Ciência da Computação, Administração, Gestão Pública e áreas afins) e abordagem conceitual, teórica ou metodológica relacionada à temática analisada de forma qualitativa, baseada no conhecimento prévio da autora. Destaca-se que foram selecionados trabalhos de áreas correlatas e não somente da CI, visto que os dados abertos têm sido estudados sob diversas perspectivas, e as buscas delimitadas à CI retornaram um número pequeno de trabalhos.

O próximo critério executado foi a **análise** qualitativa dos títulos, para definir os que tinham relação com a temática em questão e, após, foi feito o mesmo com o resumo dos documentos. Na análise qualitativa dos títulos, foram considerados todos os trabalhos que indicavam alguma abordagem possivelmente relacionada a dados abertos ou termos correlatos, como ciência aberta, governo aberto, governo eletrônico, dados abertos conectados (*linked open data*) e *big data.* A análise dos resumos deu-se de forma mais rígida, selecionando para análise completa somente os trabalhos que apresentavam como resultado alguma proposição de modelo, métrica, *framework, software* ou similares para tratamento ou avaliação dos DGA. O protocolo detalhado é apresentado no Quadro 11:

Quadro 11 - Protocolo de Revisão Integrativa de Literatura na BDTD

Protocolo de Revisão Integrativa de Literatura	
Critérios	**Descrição**
Objetivo geral	Identificar estudos sobre métricas, modelos e *frameworks* para avaliação dos dados governamentais abertos no contexto nacional

Questões a serem resolvidas	Quais trabalhos abordam métricas, modelos ou *frameworks* para a publicação, a avaliação, o tratamento ou o ciclo de vida dos DGA?
Fontes de informação pesquisadas	Base de dados: BDTD
Expressões de busca	Obs.: as expressões foram utilizadas em português, inglês e espanhol (métricas OU modelos OU *frameworks*) E (dados abertos governamentais OU dados governamentais abertos OU dados abertos OU dados governamentais OU dados públicos)
String	"(Todos os campos:"dados abertos governamentais" OR "dados governamentais abertos" OR "dados abertos" OR "dados governamentais" OR "dados públicos" OR "open data" OR "open government data" OR "public data" OR "datos abiertos del gobierno" OR "datos abiertos" OR "datos públicos abiertos" or "datos abiertos de la administración pública" OR "datos gubernamentales abiertos") E (Todos os campos:"modelo" OR "model" OR "métrica" OR "*framework*" OR "metric")"
Campos de busca	Todos os campos (Título, Autor, Assunto, Resumo Português, Resumo Inglês, Editor, Ano de Defesa)
Critérios de elegibilidade	Idioma: inglês, espanhol, português Sem delimitação de data Tipos de documentos: teses e dissertações
Critérios de inclusão e de exclusão	Inclusão: documentos nos idiomas inglês, português e espanhol; Trabalhos na área de CI ou áreas afins (Computação, Administração, Gestão Pública)
	Exclusão: exclusão de documentos que não foram escritos em português, inglês e espanhol; documentos cujos títulos e resumos estivessem fora do escopo da pesquisa; documentos duplicados
Procedimentos de seleção dos documentos recuperados	Leitura dos títulos e resumos dos documentos recuperados com o intuito de verificar a pertinência do conteúdo ao objetivo geral da pesquisa
Procedimentos de análise	Identificar métricas, modelos ou *frameworks* para a publicação, o tratamento ou a avaliação dos DGA
Critério de exclusão após análise dos documentos	Trabalhos que não possuem abordagem conceitual, teórica ou metodológica sobre métrica, modelos ou *frameworks* sobre dados governamentais abertos
Resultados	Descrição qualitativa dos principais pontos encontrados nas bibliografias selecionadas e análise de conteúdo, de acordo com a metodologia proposta por Bardin (2016).
Tratamento de dados	Análise das publicações com auxílio do organizador de referências Mendeley e Planilhas do Google Docs

Fonte: elaborado pela autora (2024).

O protocolo de revisão integrativa de literatura foi executado entre os meses de dezembro de 2023 e janeiro de 2024, na BDTD. A análise dos dados utilizou uma metodologia mista, a partir de métodos qualitativos e quantitativos. A análise dos conteúdos deu-se por meio da aplicação simplificada da metodologia descrita por Bardin (2016), seguindo as fases de pré-análise, exploração do material e tratamento dos resultados.

Os trabalhos foram analisados inicialmente em relação ao momento do tratamento dos dados (anteriores ou posteriores à publicação) e depois agrupados em duas categorias principais: (1) Tipologias dos dados abertos investigados nos estudos e (2) Temáticas relacionadas à CI, identificadas de maneira qualitativa.

Nesta etapa, também mapearam-se alguns trabalhos correlatos, com proposições que possuem similaridades e contrastes com esta dissertação, que instigaram o pensamento crítico em vários pontos de desenvolvimento desta proposta, sendo eles: Santana (2013), Rocha (2016), Pereira (2017), Martins (2018) e Penteado (2020), conforme apresentado na seção 5.1.1.

4.2.2 Etapa 2: mapeamento do contexto brasileiro

A realização desta etapa correspondeu ao cumprimento do segundo objetivo específico da pesquisa: "**Mapear documentos e legislações (leis, decretos, normativas, portarias, resoluções e instrumentos) sobre os dados abertos no governo federal brasileiro**" e forneceu subsídio para a subseção 5.2.1, que apresenta o contexto histórico dos DGA. Os procedimentos realizados são apresentados a seguir.

Inicialmente, foi realizada uma pesquisa documental, desenvolvida no segundo semestre de 2024, para identificar as principais legislações já mapeadas e publicadas pela literatura da área, e observados os instrumentos da Política de Dados Abertos do Brasil, a saber: o Painel de Monitoramento de Dados Abertos do Governo Federal e o Portal Brasileiro de Dados Abertos, conforme apontado por Nascimento Silva (2022).

Esse mapeamento iniciou-se por um levantamento prévio realizado no contexto do projeto de pesquisa "Observatório de Dados Abertos", disponível no endereço eletrônico: https://observatoriodedadosabertos.eci.ufmg.br/legislacoes/, com informações atualizadas até março de 2024, e partindo para a pesquisa bibliográfica e documental exploratória em fontes de informação governamental, como Planos de Dados Abertos, guias e manuais, iniciando pelos

indicados no próprio Portal Brasileiro de Dados Abertos, na seção ‘Publicações’, disponível no endereço eletrônico: https://dados.gov.br/dados/conteudo/publicacoes.

Assim, a partir dessa coleta de dados, do estudo de legislações e dos documentos disponibilizados, foram realizadas análises filtrando os elementos teóricos e práticos que se relacionavam com o tema da dissertação, sob a ótica da OI. Essa análise foi feita por meio de uma observação direta e resultou na seção 2: Dados Abertos na Ciência da Informação, que complementou o referencial teórico e as próximas etapas da pesquisa.

4.2.3 Etapa 3: mapeamento no Portal Brasileiro de Dados Abertos

A realização desta etapa correspondeu ao cumprimento do terceiro objetivo específico da pesquisa: “**Identificar e analisar as tipologias e os formatos de dados abertos publicados, os padrões e os metadados utilizados no Portal Brasileiro de Dados Abertos**” e forneceu subsídio para a seção 2, que apresenta o contexto histórico dos DGA e sua relação com a CI. Os procedimentos realizados são apresentados a seguir.

Como instrumento para mapear as tipologias de dados abertos publicados, formatos, padrões e metadados utilizados pelo governo federal brasileiro foi selecionado o Portal Brasileiro de Dados Abertos, uma vez que o Portal de Monitoramento não disponibiliza os dados, somente informações sobre o planejamento e a disponibilização das bases.

Após a seleção do instrumento, ele foi explorado a fim de identificar as tipologias de dados abertos, formatos, padrões e metadados utilizados, para detectar lacunas e pontos de melhoria. Para a identificação dos metadados utilizados nos conjuntos de dados publicados no Portal, foi realizada uma observação direta em conjuntos de dados apresentados e a identificação de metadados nos manuais para publicadores de dados, disponíveis nos endereços eletrônicos:

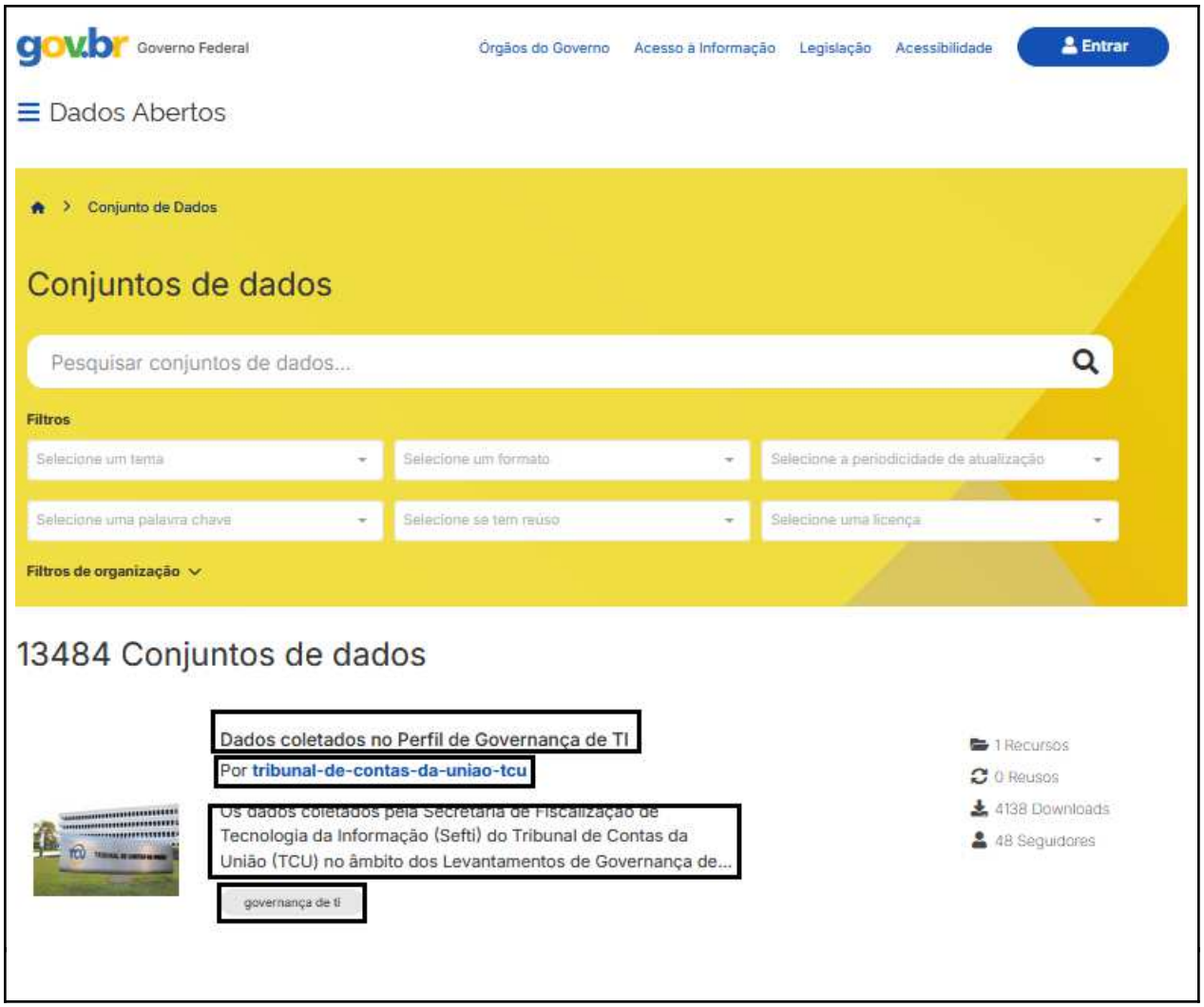
a) Como publicar um conjunto de dados: https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-conjunto-de-dados
b) Como publicar um reúso: https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-reuso
c) Cartilha Técnica para Publicação de Dados Abertos no Brasil:

https://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Cartilha_Tecnica.pdf

A observação direta consistiu na navegação contínua pelo Portal, como usuária comum, para identificar e compreender elementos da organização da informação presentes na ferramenta. Durante os meses de julho a outubro de 2024, aproximadamente, o Portal Brasileiro de Dados Abertos foi acessado diariamente por meio do endereço eletrônico: https://dados.gov.br/home e foram examinadas as opções de navegação e consulta disponíveis. Foram acessados os menus disponíveis e utilizadas as ferramentas de busca para analisar a apresentação das informações para o usuário.

Em consonância com essa etapa metodológica, os instrumentos de busca indicados na pesquisa de Oliveira, Nascimento Silva e Mafra Pereira (2024) foram utilizados para a observação direta, a saber: busca simples, busca por categorias e filtros diversos. Primeiramente, foi selecionada a opção ‘Conjuntos de Dados’, na qual são abertas várias opções de filtros para consulta e uma listagem de conjuntos de dados, em que não há especificação sobre o critério de ordenação. Nessa página já foi possível identificar alguns metadados, ainda que a maioria não possuísse rótulos, como, por exemplo, o Título do conjunto de dados, a Descrição do conjunto de dados, a Organização responsável e Palavras-chave, como exibido na Figura 4:

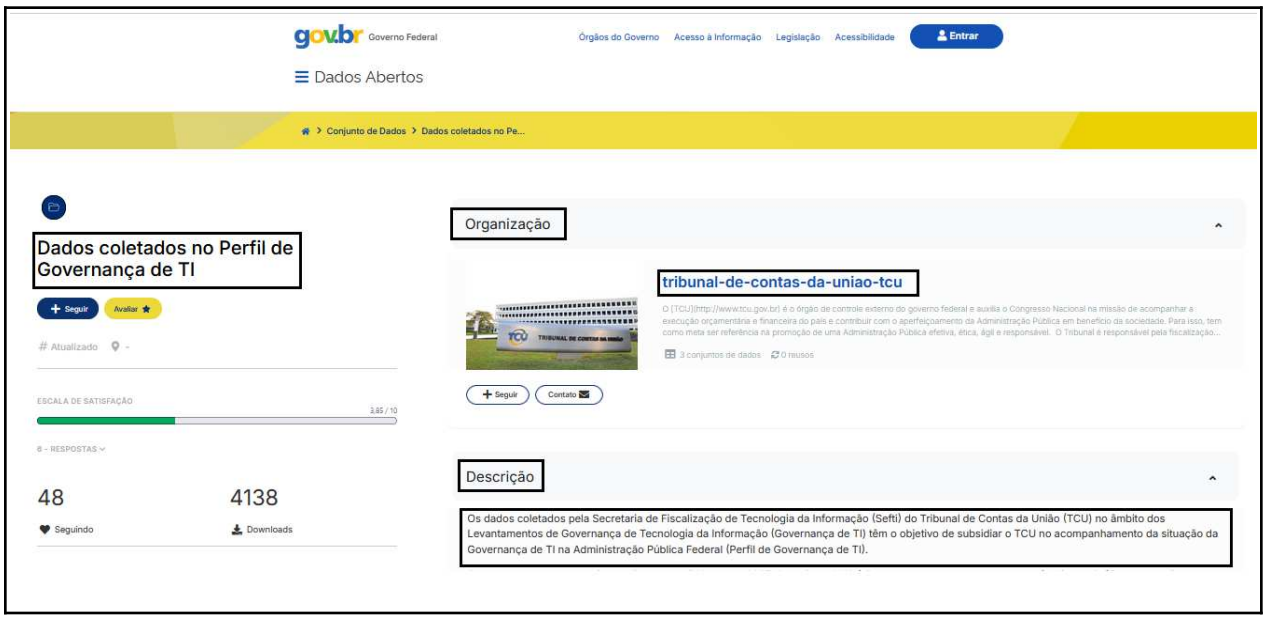
Figura 4 - Metadados no Portal Brasileiro de Dados Abertos



Fonte: Brasil (s.d.c, *online,* grifo nosso).

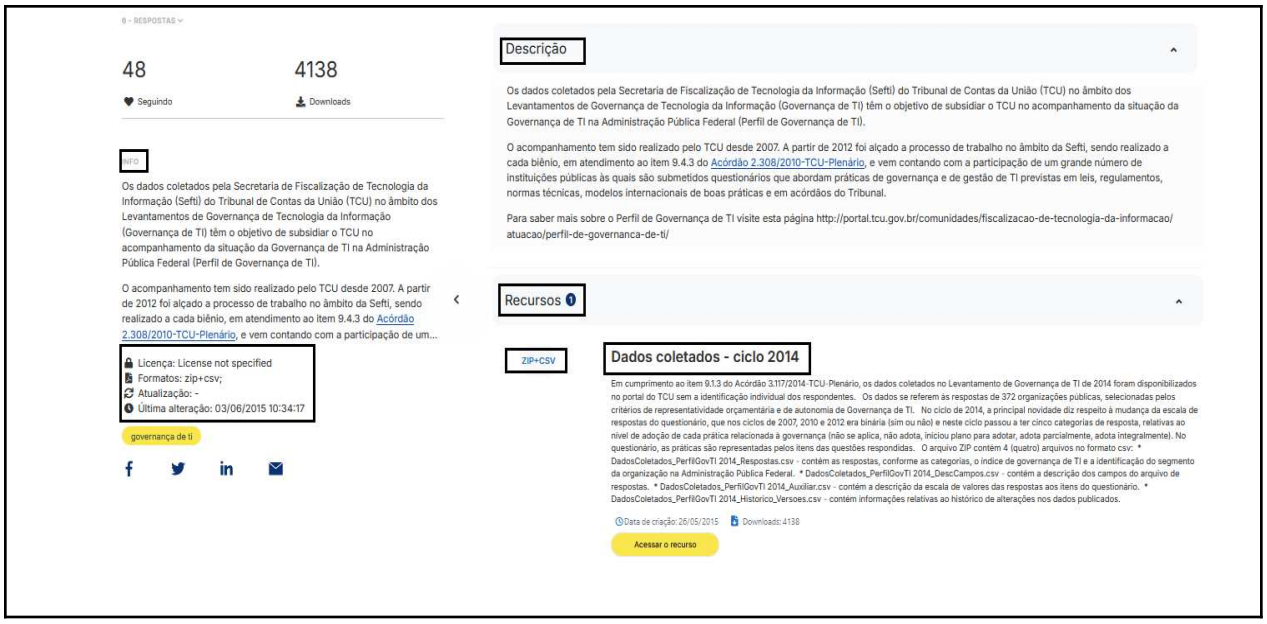
Clicou-se nos primeiros conjuntos de dados exibidos, nos quais foi possível identificar outros metadados e elementos da estrutura de organização do Portal, como, por exemplo, Dados da Organização e características dos Conjuntos de Dados e dos Recursos, como demonstrado nas Figuras 5 e 6:

Figura 5 - Metadados de um Conjunto de Dados



Fonte: Brasil (s.d.c, *online*).

Figura 6 - Metadados de um Conjunto de Dados



Fonte: Brasil (s.d.c, *online,* grifo nosso).

A identificação dos padrões de metadados para a representação de recursos informacionais utilizados emergiu juntamente com a pesquisa bibliográfica e documental da Etapa 2, na qual surgiu, inicialmente, o padrão e-PMG, mapeado na subseção 3.3.2, e, posteriormente, o DCAT-BR. Como o e-PMG utiliza como base o padrão DC e o DCAT-BR utiliza o DCAT, esses dois padrões foram mapeados e apresentados nas subseções 3.3.3 e 3.3.4, respectivamente. Por fim, apresentou-se o padrão da Cartilha Técnica para Publicação de Dados Abertos na subseção 5.2.3.

Sendo o e-PMG o instrumento oficial de representação do governo federal, realizou-se uma comparação dos metadados mapeados no Portal com os metadados apresentados no e-PMG, exibida também na subseção 5.2.3. Esse padrão também foi considerado para as discussões da Etapa 4, relatada na subseção seguinte.

4.2.3 Etapa 4: proposição de metadados para representação dos dados abertos do governo federal

Esta etapa correspondeu à realização do quarto objetivo específico da pesquisa: "**Propor uma estrutura de representação descritiva para os dados abertos do governo federal brasileiro no Portal Brasileiro de Dados Abertos com a utilização de metadados**" e consistiu na sistematização das três primeiras etapas para a consolidação da proposta.

Após o mapeamento do contexto brasileiro, foram analisados os padrões de metadados compatíveis e selecionado o padrão DCAT-3 para realizar as proposições de representação dos dados abertos do governo federal no Portal Brasileiro de Dados Abertos (instrumento oficial de disponibilização de dados). Este padrão foi selecionado por estar alinhado com as melhores práticas para dados na web, sendo, inclusive, uma das recomendações do W3C, por promover a sustentabilidade ao utilizar-se de elementos advindos de outros vocabulários (Tomoyose, 2021) e por ser um padrão específico para a descrição de catálogos governamentais. A escolha desse padrão satisfaz também as diretrizes do e-PING por ser um vocabulário recomendado pelo W3C e utilizar a sintaxe RDF.

Existe uma sinalização de que esse padrão será adotado como base para estabelecer o novo modelo de catalogação dos dados abertos do governo federal, como explicitado na subseção 5.2.2, segundo o Plano de Dados Abertos da CGU de 2024 (CGU, 2024), tornando sua utilização condizente com o cenário atual. Contudo, até a conclusão dessa dissertação, em novembro de 2024, nenhuma publicação oficial foi realizada.

Para a elaboração da proposta, primeiramente, foi realizado um estudo detalhado do padrão DCAT-3, incluindo uma análise das diretrizes e das especificações técnicas que orientam sua utilização. Em seguida, foram examinadas aplicações reais do DCAT, objetivando identificar as práticas mais recorrentes e compreender a implementação desse padrão em contextos semelhantes. Foram estudados alguns perfis DCAT, sugeridos pela própria documentação, sendo eles:

- The DCAT application profile for data portals in Europe (https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe);
- Profilo metadatazione DCAT-AP_IT (https://docs.italia.it/italia/daf/linee-guida-cataloghi-dati-dcat-ap-it/it/stabile/dcat-ap_it.html);
- Linking data portals across Belgium (http://dcat.be/);

- Clarifications, translations and explanations of DCAT-AP for Sweden (https://docs.dataportal.se/dcat/en/).

Os perfis disponibilizados foram acessados para a observação da utilização do padrão, pautando-se principalmente na avaliação de esquemas visuais, diminuindo a curva de aprendizado, tendo em vista o fato de que as documentações são extensas e complexas e demandariam uma grande carga de leitura.

Com base nas análises das Etapas 1, 2 e 3, observando as lacunas encontradas na revisão de literatura e as boas práticas para a publicação de dados na web, procedeu-se à organização hierárquica da estrutura de metadados observada no Portal Brasileiro de Dados Abertos, visando a uma sistematização preliminar dos elementos a serem modelados. Posteriormente, foi proposta uma estrutura de metadados, formulada a partir do conhecimento adquirido durante o desenvolvimento da pesquisa e com base nas necessidades específicas identificadas na observação direta para os dados abertos do governo brasileiro.

Uma preocupação surgida durante as etapas anteriores foi a adoção de boas práticas para a representação e a disponibilização de dados na web e, por isso, estudou-se com mais profundidade uma recomendação muito difundida e utilizada no cenário trabalhado para apresentar uma proposta consistente e pertinente à discussão apresentada: as Boas Práticas para Dados na Web (*Data on the Web Best Practices*), que foi apresentada na subseção 3.5.

Após a estruturação conceitual da proposta, foram estabelecidas correspondências entre os campos de metadados propostos e as propriedades do padrão DCAT-3 e, por extensão, dos padrões DC, Foaf, vCard e Adms, que, como explicitado na seção 3, são interoperáveis e complementares ao DCAT-3. Com a utilização do módulo gratuito da ferramenta Lucid Chart[1], foi criado um diagrama de classes para apresentar os relacionamentos da estrutura proposta.

Para a validação desta estrutura, ela foi transcrita para sintaxe Turtle, que serve para expressar dados em RDF, e testada na ferramenta gratuita *IDLab Turtle Validator*[2]. Essa sintaxe foi apresentada de forma completa no Apêndice A. A proposta foi elaborada entre os meses de agosto e novembro de 2024. Na próxima seção, serão apresentados os resultados obtidos na pesquisa.

5 PANORAMA DE DADOS ABERTOS NO BRASIL

Nesta seção, são descritos os resultados obtidos nas três primeiras etapas da pesquisa, oferecendo uma análise detalhada de cada uma delas.

5.1 Revisão integrativa de literatura

Esta subseção apresenta o resultado da revisão integrativa de literatura realizada com o intuito de **identificar na literatura acadêmica brasileira modelos, métricas, *frameworks* e instrumentos similares que abordem a publicação, a avaliação, o ciclo de vida ou o tratamento de dados abertos ou DGA**, conforme primeiro objetivo específico deste estudo. O protocolo criado para essa revisão de literatura bem como a metodologia utilizada foi detalhado na subseção 4.2.1, na seção de metodologia desta dissertação.

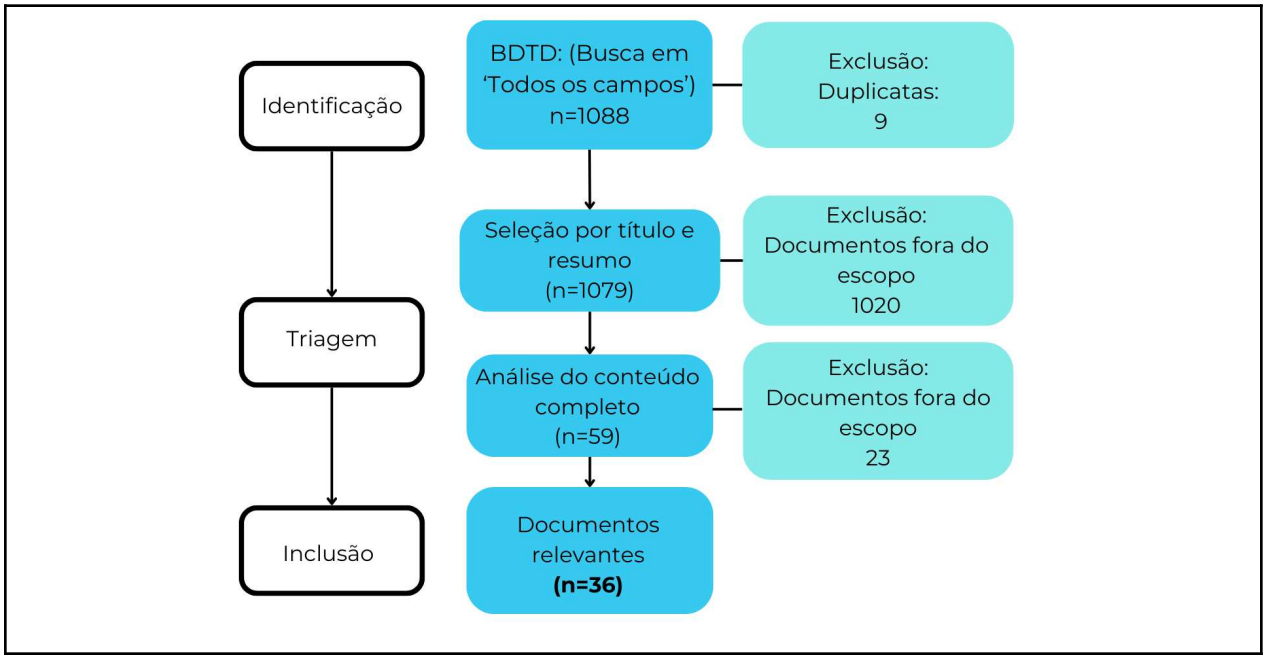
Na execução do protocolo de revisão dessa pesquisa foram obtidos 1.088 resultados, sendo 1.042 em português, 39 em inglês e um em espanhol. Em relação à tipologia documental, foram recuperadas 712 dissertações e 376 teses, datadas de 1905 a 2023. A base só permitia a exportação em formato CSV de 1.000 resultados por vez, sendo assim optou-se por fracionar a exportação dos resultados em duas vezes, a primeira considerando os trabalhos de 1905 até 2018, com 638 resultados e a segunda com os trabalhos datados de 2019 a 2023 com 450 resultados. As planilhas em CSV foram importadas para o *software* Google Planilhas agrupadas e ordenadas primeiro por título e depois por data.

Conforme os procedimentos de análise definidos no protocolo, foram excluídos nove resultados duplicados e 1.079 trabalhos foram submetidos à análise qualitativa de título. Destes, 966 trabalhos foram excluídos por não se relacionarem à questão de pesquisa, como, por exemplo: trabalhos que utilizavam os DGA para alguma análise, porém não tratavam diretamente sobre eles ou que não apresentavam um resultado que pudesse ser replicado a algum conjunto dos DGA ou trabalhos que possuíam alguma relação com termos correlatos, mas não com os DGA diretamente.

Após essa etapa, foram selecionados e submetidos à análise qualitativa de resumo 113 trabalhos, que resultaram em 59 trabalhos selecionados para a análise completa, correspondendo a fase de pré-análise do material. Durante a análise qualitativa dos 59 trabalhos selecionados, foram excluídos 23 por não abordarem especificamente DGA ou os utilizarem em outros contextos como: questões relativas à inovação ou a percepção dos publicadores (usuários), sobre

instruções recebidas, correspondendo à fase de exploração do material. Sendo assim, foram selecionados **36 trabalhos** para apresentação, discussão e execução da terceira fase de análise, que foi o tratamento dos resultados. A Figura 7 apresenta um resumo das etapas e a aplicação dos critérios de inclusão e exclusão da revisão de literatura:

Figura 7 - Etapas da seleção de documentos



Fonte: elaborada pela autora (2024).

Os trabalhos selecionados foram organizados e explanados no Quadro 12, contendo título, autor, ano, instituição (sigla), área (Programa), tipologia documental (T para teses de doutorado e D para dissertações de mestrado), momento de tratamento dos dados ou de avaliação, tipologias dos dados abertos investigados e temáticas relacionadas à CI.

Quadro 12 - Documentos da BDTD selecionados para análise

ID	Título	Autor	Ano	Instituição	Área	Tipo	Momento de tratamento dos dados ou de avaliação	Tipologias dos dados abertos investigados	Temáticas relacionadas à CI
T1	*A framework for exploiting open data to improve spatial keyword query applications*	ALMEIDA, J. P. D.	2021	UFBA	Doutorado em Ciência da Computação	T	Pós-publicação	Dados abertos geoespaciais	Recuperação de informação
T2	Dados abertos governamentais: uma análise aplicada ao Ministério Público do Trabalho	PAULO, J. S.	2018	UFES	Mestrado em Gestão Pública	D	Pré-publicação	DGA do Ministério Público	Gestão de informação
T3	Dados abertos governamentais: uma proposta de classificação e estruturação para abertura dos dados de IFES	ARRUDA, W. S.	2019	UFRPE	Mestrado Profissional em Administração Pública	D	Pré-publicação	DGA de instituições de ensino	Representação descritiva de informação, gestão de informação
T4	Dados governamentais abertos: métricas e indicadores de reúso	SILVA, P. N.	2018	UFMG	Doutorado em Gestão e Organização do Conhecimento	T	Pós-publicação	DGA	Gestão de informação
T5	Dados governamentais abertos: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor	ALBANO, C. S.	2014	USP	Doutorado em Administração	T	Pré e pós-publicação	DGA	Gestão de informação
T6	Desafios e possibilidades para o acesso às informações públicas e a transparência ativa na Universidade Federal de Juiz de Fora – MG	ANDRADE, T. S.	2020	UFJF	Mestrado em Gestão e Avaliação da Educação Pública	D	Pré-publicação	DGA de instituições de ensino	Gestão de informação, visualização de informação

T7	Desenvolvimento da interface e automatização da coleta de dados na *Open Smart City View*	RIGO, F.	2018	UPF	Mestrado em Computação Aplicada	D	Pós-publicação	DGA	Organização e tratamento de informação, visualização de informação
T8	Do mapeamento de dados abertos à modelagem conceitual: um *Data Warehouse* para análise de informações multidimensionais	PRECHT P. A. C.	2019	UFRGS	Mestrado em Computação	D	Pré e Pós-publicação	DGA	Visualização de informação
T9	Estudo sobre aplicação de *web* semântica e visualização em dados abertos: proposta de portal para transparência em séries históricas de indicadores sociais, econômicos e financeiros, públicos na *web*	GARCIA, P. S. R.	2011	PUC_SP	Mestrado em Tecnologias da Inteligência e Design Digital	D	Pós-publicação	DGA orçamentários	*Web* semântica, visualização de informação
T10	Governança de dados abertos governamentais: *framework* conceitual para as universidades federais, baseado em uma visão sistêmica	CASAES, J. C. C.	2019	UFSC	Doutorado em Engenharia e Gestão do Conhecimento	T	Pré-publicação	DGA de instituições de ensino	Gestão de informação, governança de dados
T11	*Government institutions and their documents: an ontological approach*	PINTO, J. A.	2020	UFMG	Doutorado em Gestão e Organização do Conhecimento	T	Pré-publicação	DGA	Representação temática de informação, *web* semântica
T12	IDEO Integrador de dados da Execução Orçamentária Brasileira: um estudo de caso da integração de dados das receitas e despesas nas Esferas Federal, Estadual Governo de São Paulo, e Municipal Municípios do Estado de São Paulo	BELUZO, J. R.	2015	USP	Mestrado em Ciências	D	Pós-publicação	DGA orçamentários	Recuperação e visualização de informação

T13	Integração, visualização e análise de informações eleitorais usando bancos de dados analíticos e fontes heterogêneas de grande volume	PIRES, L. B.	2015	UFABC	Mestrado em Ciência da Computação	D	Pós-publicação	DGA eleitorais	Visualização de informação
T14	LigADOS: interligação de conjuntos de dados em plataformas de dados abertos na *web* semântica	FIGUEIREDO, G. B.	2020	UFRJ	Mestrado em Informática	D	Pós-publicação	Dados abertos	Dados conectados, *web* semântica
T15	Modelagem e representação semântica de dados governamentais abertos da Previdência Social brasileira	PEREIRA, D. V.	2014	UFF	Mestrado em Ciência da Informação	D	Pré-publicação	DGA previdenciários	Representação temática de informação, *web* semântica, dados conectados
T16	Modelo de apoio à decisão para avaliar dados governamentais abertos do setor elétrico	ARAÚJO, I. P.	2022	UNB	Mestrado Profissional em Computação Aplicada	D	Pré-publicação	DGA do setor elétrico	Gestão de informação, visualização de informação
T17	Modelo de infraestrutura para publicação de dados abertos governamentais conectados de qualidade	PENTEADO, B. E. P.	2020	USP	Doutorado em Ciências	T	Pré-publicação	DGA	Dados conectados
T18	Modelo de maturidade de dados abertos: uma matriz de referência para organizações	VISINTIN, L.	2021	UFSC	Doutorado em Engenharia e Gestão do Conhecimento	T	Pré e pós-publicação	Dados abertos	Gestão de informação
T19	Modelo de referência para indicadores de inovação regional suportado por dados ligados	SPERONI, R. M.	2016	UFSC	Doutorado em Engenharia e Gestão do Conhecimento	T	Pré-publicação	Dados abertos	Dados conectados
T20	OGDPub: uma ontologia para publicação de dados abertos governamentais	PEREIRA, L. M. F.	2017	UFSC	Mestrado em Engenharia e Gestão do Conhecimento	D	Pré-publicação	DGA	Representação temática de informação

T21	Ontologias como artefatos para a representação de dados abertos de pesquisa: uma exploração acerca de uma modelagem experimental de dados da Arqueologia através do modelo CIDOC *Conceptual Reference Model*	OLIVEIRA, C. C.	2023	UFRGS	Mestrado em Ciência da Informação	D	Pré e pós-publicação	Dados abertos de pesquisa	Representação temática de informação
T22	Ontologias como um agente para a interligação de dados abertos em contexto governamental: características e aplicabilidade	ROCHA, L. L.	2016	UFF	Mestrado em Ciência da Informação	D	Pós-publicação	DGA	Representação temática de informação, dados conectados
T23	*Open smart city view*: uma solução computacional para manipulação e apresentação de dados governamentais	LUSA, D. A.	2016	UPF	Mestrado em Computação Aplicada	D	Pós-publicação	DGA	Organização e tratamento de informação, visualização de informação
T24	Plano de dados abertos: estudo multicaso para a priorização de dados a partir da transparência passiva na Universidade Federal de Santa Catarina	CUNHA, P.	2020	UFSC	Mestrado em Administração Universitária	D	Pré-publicação	DGA de instituições de ensino	Gestão de informação
T25	Proposta de arquitetura de publicação automatizada de dados abertos conectados utilizando meta-dados e ontologias	MARTINS, L. C. B.	2018	UNB	Mestrado em Computação Aplicada	D	Pré-publicação	Dados abertos	Representação temática de informação
T26	Publicação de dados conectados sobre despesas orçamentárias do governo federal brasileiro	FANTINI, W. S.	2015	UFPE	Mestrado em Ciência da Computação	D	Pré-publicação	DGA orçamentários	Representação temática de informação, dados conectados
T27	Publicação de dados de pesquisa científica: proposta de estruturação semântica de cadernos abertos de pesquisa frente às dimensões da *e-Science*	SILVA, L. C.	2020	UNESP	Doutorado em Ciência da Informação	T	Pré-publicação	Dados abertos de pesquisa	*Web* semântica

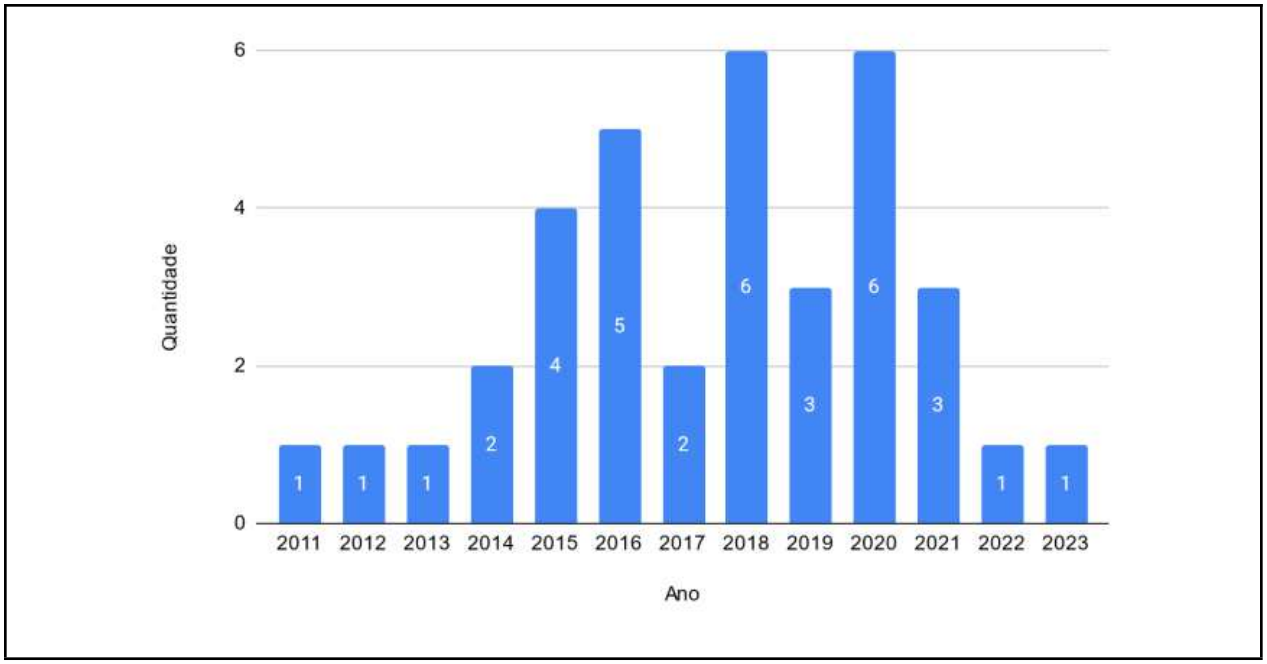
T28	Transparência ativa para acesso à informação: uma proposta de modelo com os temas prioritários em Instituições Federais de Ensino Superior	SOUZA, D. F.	2018	UFRPE	Mestrado em Administração Pública	D	Pré-publicação	DGA de instituições de ensino	Gestão de informação
T29	Um modelo de avaliação de conformidade de portais de dados abertos de governo	GOMES, M. S.	2021	UFSC	Doutorado em Engenharia e Gestão do Conhecimento	T	Pós-publicação	DGA	Gestão de informação
T30	Um modelo para integração de informações de bases de dados abertos, com uso de ontologias	TOSIN, T. M.	2016	UNISINOS	Mestrado em Computação Aplicada	D	Pós-publicação	Dados abertos	Representação temática de informação, dados conectados
T31	Um processo para publicação de dados abertos em institutos federais baseado em BPM	LEMOS, J. M. M.	2017	UFPE	Mestrado em Ciência da Computação	D	Pré-publicação	DGA de instituições de ensino	Gestão de informação
T32	Uma abordagem para a publicação de dados abertos ligados obtidos a partir de bases de dados relacionais	PEREIRA, C. M.	2012	ITA	Mestrado em Engenharia Eletrônica e Computação	D	Pré-publicação	Dados abertos	*Web* semântica, Representação temático de informação, visualização de informação, dados conectados
T33	Uma plataforma para disponibilização centralizada de dados abertos governamentais como suporte para aplicações no contexto de cidades inteligentes.	VEIRA, D. I.	2016	UFSCAR	Mestrado em Ciência da Computação	D	Pós-publicação	DGA	Organização e tratamento de informação, representação descritiva de informação
T34	Uma proposta de modelo de processo para publicação de dados abertos conectados governamentais	ÁVILA, T. J. T.	2015	UFAL	Mestrado em Modelagem Computacional do Conhecimento	D	Pré-publicação	DGA	Dados conectados

T35	Uma proposta de processo para implantação de dados abertos em instituições públicas brasileiras	SILVA, R. O.	2018	UFRN	Mestrado em Sistemas e Computação	D	Pré-publicação	DGA de instituições de ensino	Gestão de informação
T36	Uma proposta de publicação de dados do orçamento público na *Web*	SANTANA, M. T.	2013	USP	Mestrado em Ciências	D	Pré-publicação	DGA orçamentários	Representação temática de informação

Fonte: elaborado pela autora (2024).

Dos 36 documentos selecionados para discussão, 34 (94,4%) estavam no idioma português e dois (5,6%) estavam em inglês. Em relação à tipologia documental, foram selecionadas 26 dissertações de mestrado (72,2%) e 10 teses de doutorado (27,8%). Os trabalhos foram publicados por 21 instituições de ensino, com destaque para a Universidade Federal de Santa Catarina (UFSC), responsável por seis (16,7%) das publicações e a Universidade de São Paulo (USP), responsável por quatro (11,1%) delas. Os trabalhos foram datados entre os anos de 2011 e 2023, conforme mostrado no Gráfico 1.

Gráfico 1 - Distribuição de publicações por ano



Fonte: elaborado pela autora (2024).

Nessa seção, apresenta-se um panorama dos 36 trabalhos selecionados, por meio de uma caracterização descritiva, no qual foram evidenciadas algumas similaridades entre eles, realizando um agrupamento de acordo com a temática abordada em cada um, apontando semelhanças e diferenças entre as abordagens propostas.

Destaca-se que foram considerados trabalhos que desenvolveram produtos, modelos, propostas, diretrizes, protótipos, arquiteturas, *softwares*, *frameworks* ou ferramentas aplicadas ao contexto de dados abertos ou DGA. É importante destacar que alguns trabalhos recuperados contemplavam o tratamento, a organização ou a avaliação de dados abertos, contudo, por

trazerem uma contribuição relevante e possível de ser aplicada nos DGA, foram considerados na amostra analisada. A análise dos conteúdos considerou: o momento de tratamento dos dados ou a avaliação dos modelos, as tipologias de dados abertos investigados e os estudos em si, observando similaridades e distinções dos artefatos e resultados apresentados.

Observou-se que todos os trabalhos selecionados estão inseridos no período dos últimos 13 anos (2011-2023), em consonância com a evolução histórica dos dados abertos e dos DGA, sendo que em 2011 foi promulgada a LAI no Brasil e firmada a parceria *Open Government Partnership (OGP)*, o que pode ter motivado o interesse científico pelo assunto. Dentre eles, 28 trabalhos tratavam especificamente dos DGA e oito referiam-se a dados abertos de modo geral. Os trabalhos selecionados estavam concentrados majoritariamente nas áreas de Computação, Administração, Gestão Pública e Gestão do Conhecimento (35 trabalhos, 97,5%). A maior parte realizou proposições de modelos, seguidas por *softwares*.

Em uma primeira categorização, considerando o **momento de tratamento dos dados ou a avaliação dos modelos trabalhados,** foi observado que 20 trabalhos selecionados eram focados nas etapas anteriores a publicação dos dados, preocupando-se com questões como a representação desses dados, ou na etapa de publicação, 12 trataram aspectos relativos a dados já publicados, com foco no tratamento ou na recuperação de dados, por exemplo, e quatro abordaram etapas anteriores e posteriores à publicação. Dentre esses, geraram artefatos que orientam o processo de publicação dos dados: T32 - Pereira (2012), T36 - Santana (2013), T15 -Pereira (2014), T34 - Ávila (2015), T26 - Fantini (2015), T31 - Lemos (2017), T20 - Pereira (2017), T25 - Martins (2018), T35 - Silva, R. O. (2018), T24 - Cunha (2020), T17 - Penteado (2020), T27 - Silva (2020).

Nesse quesito, o presente trabalho aproxima-se daqueles que focaram nas etapas anteriores à publicação dos dados, visto que a proposta baseia-se na representação dos DGA com a utilização de metadados.

Considerando as **tipologias dos dados abertos investigados nos estudos**, vários trabalhos se ativeram a tipos específicos de DGA, como dados geoespaciais, dados da previdência social, dados do setor elétrico, dados eleitorais, etc, o que pode acarretar em dificuldades para replicar as propostas em contextos mais genéricos, entretanto, pode complementar trabalhos como a desta dissertação, que abordam os DGA em geral, mas limitam-se ao tratar superficialmente as especificidades de cada tipo de DGA.

O conjunto de dados mais abordado foi o de instituições de ensino superior, elencados em sete trabalhos, sendo eles: T3 - Arruda (2019), T6 - Andrade (2020), T10 - Casaes (2019), T24 - Cunha (2020), T28 - Souza (2018), T31 - Lemos (2017) e T35 - Silva, R. O. (2018) e observa-se que todos eles abordaram os DGA antes do momento da publicação, realizando proposições, criação de Plano de Dados Abertos (PDA) ou para estruturação dos dados para publicação, por exemplo.

Destacam-se também os trabalhos de T9 - Garcia (2011), com a proposição de um modelo de portal de transparência, T12 - Beluzo (2015), com a proposição de um integrador de de dados da execução orçamentária brasileira, T26 - Fantini (2015), abordando a publicação de dados conectados sobre despesas orçamentárias do governo federal brasileiro, e T36 - Santana (2013), com uma proposta de publicação de dados do orçamento público na Web, que somaram quatro trabalhos, abordando dados orçamentários e podendo auxiliar no alcance de princípios básicos dos DGA, como apontado por Visintin ao dizer que: “a abertura dos dados por parte de organizações públicas ou privadas é necessária e importante para que se obtenha transparência, *accountability*, participação social e colaboração” (Visintin, 2021, p. 180). Nesse caso, dois autores trabalham aspectos pré-publicação (T26 e T36) e outros dois apresentaram dados já publicados (T9 e T12).

Com relação aos **conteúdos abordados**, uma temática similar aproxima os trabalhos de T12 - Beluzo (2015), T13 - Pires (2015) e T30 - Tosin (2016). Os três estudos abordam a **integração de dados**. T13 - Pires (2015) desenvolveu um *Data Warehouse* para integração de dados eleitorais e T30 - Tosin (2016), utiliza de ontologias para realizar a integração de bases de dados abertos e conectados, a partir de dados de compras governamentais para testes do protótipo desenvolvido. Assim como T13 - Pires (2015), T8 - Precht (2019), desenvolveu um trabalho utilizando um *Data Warehouse*. Nesse caso, o autor utilizou processos de extração, transformação e carga (*Extract, Transform, Load*), para agregar dados de diversas fontes e integrá-los no *Data Warehouse.* Além destes dois, T12 - Beluzo (2015), T23 - Lusa (2016) e T7 - Rigo (2018) também aplicaram *Data Warehouse* em suas pesquisas.

Ao observar os **artefatos gerados,** bem como os resultados apresentados nos estudos, T12 - Beluzo (2015), teve como objetivo a criação de um repositório de informações de fontes de dados heterogêneas referentes à execução orçamentária nas três esferas governamentais no Brasil, em um repositório de dados integrado na arquitetura de *Data Warehouse* e inclusive lista o

trabalho de T36 - Santana (2013) como similar ao seu. Este segundo propôs-se a desenvolver um padrão de publicação de DGA da execução do orçamento público, construindo para isso uma taxonomia, que segundo T12 - Beluzo (2015), facilita o processo de integração e ligação dos dados, mas tem a limitação de depender da disponibilidade dos governos de se adequarem para publicar os dados. Ainda que uma década separe o de Santana (2013) da presente dissertação, a limitação apontada por Beluzo (2015) ainda é uma realidade.

Em consonância com os dois trabalhos citados acima, T26 - Fantini (2015), trabalha com um tipo de vocabulário controlado, assim como T36 - Santana (2013), para representação de dados orçamentários ao propor uma extensão ao modelo ontológico do Orçamento Federal Brasileiro, utilizando princípios de *linked data* para geração de dados conectados.

Uma característica que atraiu a atenção, foi a de que poucos trabalhos se debruçaram sobre **avaliação dos DGA,** sendo eles: T4 - Silva, P. N. (2018), que desenvolveu uma métrica brasileira para avaliação do reúso dos DGA; T16 - Araújo (2022), com a proposição de um modelo para avaliar DGA do setor elétrico; T29 - Gomes (2021) que trouxe um modelo para avaliar a conformidade de portais dos DGA; e T18 - Visintin (2021), com um modelo para avaliação de maturidade de dados abertos.

Silva, P. N. (2018) mantém até os dias de hoje o ineditismo de seu trabalho, um modelo de avaliação dos DGA na Administração Pública Federal do Brasil, utilizando-se de métricas e indicadores internacionais para amparar a avaliação do reúso dos DGA brasileiros. A métrica é voltada para dados já publicados e tem cinco perspectivas de avaliação, estando disponível no endereço eletrônico: https://www.dgabr.com.br/.

Visintin (2021) foi o único trabalho recuperado que tratou sobre maturidade de dados, sendo que um modelo para tal “fornece uma estrutura que permite a uma organização avaliar o quão bem seus processos se adequam às melhores práticas da indústria” (Instituto de Dados Abertos, 2016).

Outros trabalhos que também geraram documentos norteadores como artefatos de suas pesquisas foram T2 - Paulo (2018), com um modelo de PDA aplicado ao Ministério Público do Trabalho; T6 - Andrade (2020), apresentando um Plano de Ação Educacional (PAE) com o objetivo de aprimorar o fluxo do acesso à informação na Universidade Federal de Juiz de Fora e a disponibilização de dados públicos no seu Portal da Transparência e T24 - Cunha (2020),

desenvolvendo um modelo de priorização de dados para publicação no formato de dados abertos pela Universidade Federal de Santa Catarina (UFSC).

Foi observado que alguns trabalhos se debruçaram sobre o estudo e análise de Portais de DGA, porém cada um deles possuía um direcionamento. O trabalho de T29 - Gomes (2021) une-se ao trabalho de T9 - Garcia (2011), por exemplo, no sentido de serem os únicos a abordarem os portais dos DGA, também conhecidos como portais da transparência, ainda que sob perspectivas diferentes. Enquanto o primeiro tratou da avaliação de portais já disponibilizados, o segundo realizou a proposição de um modelo para a construção de um portal que dê transparência a séries históricas de indicadores sociais, econômicos e financeiros, públicos e disponíveis na web.

Ainda abordando a temática de portais dos DGA, foram selecionados três trabalhos que abordaram os DGA no contexto das chamadas Cidades Inteligentes, que são aquelas que visam explorar as Tecnologias de Informação e Comunicação (TICs) mais avançadas para apoiar serviços de valor agregado para a administração da cidade e para os cidadãos (Zanella *et al.,* 2014), sendo eles: T23 - Lusa (2016), que desenvolveu um modelo tecnológico de coleta (*Open smart city view* - OSCV), processamento e apresentação de informações a partir de dados coletados em portais de transparência governamental; T7 - Rigo (2018), que apresentou melhorias para o trabalho de Lusa (2016), com o desenvolvimento um processo de automação e uma interface web simples para a ferramenta OSCV; e T33 - Vieira (2016), com um trabalho bastante similar aos dois primeiros, desenvolvendo uma plataforma para coleta, armazenamento e disponibilização centralizada dos DGA, sendo uma fonte de dados para projetos no cenário de cidades inteligentes. O trabalho de Vieira (2016), apresenta o diferencial de pensar na organização dos DGA após a coleta, tendo em uma de suas camadas um dicionário de dados que registra informações de metadados.

Em relação às **temáticas relacionadas à CI**, um dos assuntos mais abordados nos trabalhos, estava relacionado à **representação de informação**, que pode ser subdividida em representação descritiva e representação temática, conceitos abordados no referencial teórico na perspectiva de Maimone, Silveira e Tálamo (2011).

Conteúdos sobre representação temática de informação, como vocabulários controlados, ontologias, taxonomias e tesauros foram abordados por T36 - Santana (2013), T22 - Rocha (2016), T30 - Tosin (2016), T20 - Pereira (2017), T25 - Martins (2018), T11 - Pinto (2020) e T21

- Oliveira (2023). Observa-se, por outro lado, que abordagens relativas à representação descritiva dos DGA não estão tão presentes nos trabalhos recuperados, representando uma oportunidade de estudo, explorada nesta dissertação.

Outro assunto que possui relação direta com a representação e recuperação de informação e esteve presente em sete trabalhos relaciona-se aos **dados ligados/conectados**, que se referem a um conjunto de práticas recomendadas para publicação de dados estruturados na Web (W3C, 2023), e está presente nos trabalhos T32 - Pereira (2012), T34 - Ávila (2015), T26 - Fantini (2015), T22 - Rocha (2016), T19 - Speroni (2016), T14 - Figueiredo (2020) e T17 - Penteado (2020).

Os dados ligados/conectados estão relacionados ao conceito da web semântica que é "um projeto do W3C cujo objetivo é tornar as informações legíveis por máquina, por meio do desenvolvimento de tecnologias, linguagens, padrões e recomendações" (Catarino; Cervantes; Almeida, 2015) e está intimamente relacionada com **integração e interligação de dados** e é uma temática explícita em: T9 - Garcia (2011), T15 - Pereira (2014), T14 - Figueiredo (2020) e T27 - Silva (2020).

Sendo a legislação uma das premissas fundamentais para a disponibilização de dados públicos, a realização de trabalhos acadêmicos que apresentem propostas de aprimoramento nesse contexto torna-se relevante e promissora, promovendo a participação cidadã e gerando impacto social. Tais pesquisas contribuem para o fortalecimento da governança pública e para o fomento de políticas baseadas em evidências.

No entanto, uma limitação inferida nesta etapa, sobre os estudos que tratam da representação dos DGA, como modelos para PDA ou ontologias para representação temática, está na dependência das iniciativas do setor público para a efetivação de políticas, regulamentações ou legislações específicas que viabilizem a implementação das propostas. Isso inclusive pode justificar a observação de que são escassos os trabalhos a abordarem a representação descritiva de informação.

Na outra ponta, propostas que abordam dados já publicados, como métricas de avaliação ou para visualização dos dados disponibilizados, apresentam a vantagem significativa de poderem analisar e sugerir melhorias com base em dados já disponíveis, o que permite um embasamento mais sólido e prático em suas proposições. Além disso, a acessibilidade a dados

públicos previamente divulgados acarreta um facilitador para o desenvolvimento das análises e propostas, facilitando o processo de validação científica.

Entretanto, para ambos os casos, existe o risco inerente à dependência do setor público para a evolução e possível transformação de uma proposta em um produto, uma vez que, sem atualizações contínuas e respaldo governamental, muitas dessas propostas correm o risco de obsolescência. Exemplificando com a conclusão obtida por diversos autores estudados nesta etapa, sobre os desafios impostos pela falta de padronização e pela heterogeneidade dos DGA, observa-se que, ainda que existam propostas de solução não integradas, a ausência de um compromisso por parte do setor público, torna ineficaz a implementação dessas recomendações.

Por fim, observou-se que existem inúmeras possibilidades de proposições dentro do universo dos dados abertos e dos DGA e ainda assim, foi possível identificar diversas similaridades e aproximações entre os trabalhos selecionados, demonstrando tendências e inclinações para a abordagem destes. Desta forma, foi possível identificar boas práticas para tratar dos DGA e também lacunas e desafios explanados pelos autores, que representam oportunidades para trabalhos futuros sob a perspectiva de áreas variadas.

Esta revisão integrativa de literatura visou ao cumprimento do primeiro objetivo específico desta pesquisa, investigando na literatura acadêmica brasileira, trabalhos de pesquisa apresentados em dissertações e teses que desenvolveram modelos, métricas e ou *frameworks* que abordassem aspectos dos DGA.

Considerando a quantidade de trabalhos recuperados inicialmente, após a aplicação da *string* comparada com a quantidade de trabalhos selecionados, evidenciam-se a escassez de trabalhos que abordaram a temática investigada e as lacunas que podem representar oportunidades de pesquisa futuras. Em síntese, foi observada uma preocupação unívoca com desafios decorrentes da falta de padronização e heterogeneidade dos dados, ressaltando a importância destas para promover a interoperabilidade, a apropriação e a reutilização dos dados.

Esta revisão foi importante para delinear a pesquisa em andamento, de forma a preencher lacunas e amparar-se em trabalhos similares identificados. Com o panorama de modelos, métricas e *frameworks* produzidos em programas de pós-graduação *stricto sensu* brasileiros apresentado, foi possível extrair informações e critérios para as análises realizadas nas próximas etapas, focadas diretamente no objeto de pesquisa.

5.1.1 Trabalhos correlatos com a proposta da dissertação

Nesta seção, apresentam-se os trabalhos de Santana (2013), Rocha (2016), Pereira (2017), Martins (2018) e Penteado (2020), identificados durante a revisão integrativa, que foram considerados pertinentes para embasar a presente dissertação, pautados nas similaridades e nos contrastes com esta proposta.

Santana (2013) apresenta uma proposta de publicação de dados governamentais abertos relacionados à execução do orçamento público, em conformidade com a legislação brasileira que envolve dados, metadados e uma taxonomia. O autor cita como referenciais a e-PING, o VCGE e o e-PMG por fornecerem diretrizes para a interoperabilidade, organização e disponibilização de informações públicas de forma padronizada e acessível.

Para os metadados, Santana (2013) apresenta nove elementos do DC, porém não há aprofundamento sobre as escolhas. Uma observação importante destacada no estudo é que dados baixados sem um conjunto de metadados associado podem, ao serem transferidos e reutilizados dentro da sociedade civil, perder informações cruciais. Para evitar essa perda e preservar os princípios de transparência, integridade e reprodutibilidade, cada porção de dados publicada deve conter metadados permanentes. Assim, os metadados tornam-se um componente essencial para garantir o valor e a utilidade dos dados ao longo do tempo, perspectiva que corrobora a importância da temática desta dissertação.

O trabalho de Rocha (2016) apresenta diretrizes para a aplicação de ontologias na interligação de dados governamentais na web, considerando o contexto brasileiro e as especificações da LAI, evidenciando a necessidade de padrões que garantam a interoperabilidade e a acessibilidade dos dados públicos. O autor realiza uma análise do Portal Brasileiro de Dados Abertos à luz da LAI e revela a ausência de um vocabulário controlado, essencial para a organização e recuperação eficiente das informações.

Como solução, o trabalho de Rocha (2016) propõe o uso de ontologias para preencher essa lacuna, fornecendo um modelo semântico que melhora a qualidade e a consistência dos dados interligados. Ademais, para que essa interligação seja efetiva, é fundamental que os dados estejam padronizados, e um esquema de metadados claro e abrangente desempenha papel central nesse processo. Assim, indiretamente, a proposta apresentada nesta dissertação também pode auxiliar na padronização e no preenchimento das lacunas identificadas.

Pereira (2017) apresentou um trabalho diretamente relacionado e alinhado ao desta dissertação ao propor uma ontologia de domínio para descrição de *datasets* de municípios brasileiros. A autora pautou-se, entre outros, no padrão DCAT para fornecer um arcabouço de metadados e concluiu que ele é eficaz para garantir que os conjuntos de dados (*datasets)* sejam encontrados de maneira mais simples na web. Os trabalhos se diferenciam na medida em que este debruça-se mais especificamente no papel dos metadados para representação da informação e o padrão DCAT complementa o esquema proposto, enquanto Pereira (2017) faz suas proposições baseado no que existe no padrão e não aprofunda a discussão sobre as escolhas específicas de metadados.

Martins (2018), em sua proposição de uma arquitetura para dar suporte à publicação automatizada de dados abertos, aborda a indexação semântica, metadados, vocabulários controlados e ontologias. Um dos objetivos do autor é estabelecer um conjunto mínimo de metadados descritivos para dados abertos e, para isso, ele utiliza alguns metadados listados na Cartilha Técnica para Publicação de Dados Abertos e complementa seu conjunto com alguns elementos do DC, porém o autor não detalha essas escolhas e mantém seu foco em realizar enriquecimento semântico dos dados para torná-los dados conectados.

Por fim, Penteado (2020) apresenta o trabalho possivelmente mais similar a esta dissertação, pois um dos produtos de sua tese é um modelo genérico de metadados para dados conectados capaz de atender às necessidades de descrição de catálogos e conjuntos de dados e utiliza a estrutura do padrão DCAT para apresentá-lo. A abordagem do autor é guiada pelas 35 Boas Práticas para Dados na Web, que também foram utilizadas na presente dissertação.

Penteado (2020), ressalta a importância dos metadados no contexto dos DGA conectados, atribuindo a eles a responsabilidade de alinhar, conectar e reaproveitar fontes de dados heterogêneas, permitindo o processamento automático por agentes de *software*. Os trabalhos divergem-se no ponto em que Penteado (2020) define um conjunto mínimo de metadados comuns a diferentes vocabulários e esta dissertação aprofunda-se mais ao trabalhar com um domínio específico. O autor também cita a Cartilha Técnica para Publicação de Dados Abertos, reforçando a importância de diretrizes nacionais no contexto da publicação de dados abertos. Uma limitação apresentada é referente ao preenchimento manual de metadados, que se mostrou trabalhoso e propenso a erros, lacuna que pode ser parcialmente solucionada com as proposições de campos com valores pré-definidos e controle de vocabulário apresentadas nesta dissertação.

Observando as abordagens e soluções apresentadas nos trabalhos correlatos, vê-se que o tema de estudo escolhido é vasto, diversificado e complexo, o que torna as investigações nessa área ainda mais pertinentes e fundamentais. Ainda que os desafios apresentados sejam árduos, o estudo contínuo, a revisão das práticas adotadas e o desenvolvimento de novas propostas são a trajetória mais certeira para a obtenção do êxito.

5.2 Dados abertos do governo federal brasileiro

Essa seção apresenta os resultados obtidos em relação ao cumprimento dos segundo e terceiro objetivos específicos: "**Mapear documentos e legislações sobre os dados abertos no governo federal do brasileiro**" e "**Identificar e analisar as tipologias e os formatos de dados abertos publicados, padrões e metadados utilizados no Portal Brasileiro de Dados Abertos**", realizados a partir de pesquisa bibliográfica e documental e observação direta.

5.2.1 Iniciativas de dados abertos no governo federal brasileiro

Nessa etapa, foi realizado um mapeamento simplificado de legislações, normas, diretrizes, manuais, instrumentos e similares que compõem as iniciativas de dados abertos no governo federal brasileiro e apresentam alguma contribuição direta para a representação descritiva desses dados, traçando o histórico no contexto brasileiro com os principais marcos. Inicialmente, foi apresentado um panorama sob a ótica do programa de Governo Eletrônico, conforme publicado em Brasil (2024).

A Constituição Federal, em seu Art. 5º, inciso XXXIII, prevê o direito de todo cidadão "receber dos órgãos públicos informações interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado" (Brasil, 1988). Essa realidade foi fortalecida com o surgimento do programa de Governo Eletrônico, em 2000, e a seguir se apresenta, de forma geral, os principais marcos que se sucederam entre os anos de 2000 e 2024.

O programa de Governo Eletrônico buscava transformar as relações do governo com cidadãos, empresas e órgãos ligados a ele, aprimorando a qualidade dos serviços prestados e fortalecendo a participação cidadã por meio do acesso à informação. O Governo Eletrônico (eGov) abarca a definição de padrões, a normatização, a articulação da integração de serviços

eletrônicos, a disponibilização de boas práticas, a criação e construção de super infraestrutura tecnológicas etc (Brasil, 2024, *online*). Fruto dessa articulação, surgiu o programa "Sociedade da Informação", conduzido pelo Ministério de Ciência e Tecnologia (Socinfo/MCT), com o objetivo de coordenar ações, visando, entre outras finalidades, a ampliação do acesso da população aos benefícios gerados pelas TICs (Brasil, 2024, *online*).

No ano 2004, foi publicada a primeira versão do documento: Padrões de Interoperabilidade em Governo Eletrônico (e-PING), e, em 2005, foi publicada a Portaria Normativa nº 05, de 14 de julho, institucionalizando esses padrões no âmbito do Sistema de Administração dos Recursos de Tecnologia da Informação (SISP). Além disso, foi lançado o Modelo de Acessibilidade de Governo Eletrônico (e-MAG), recomendando a acessibilidade nos portais e sítios eletrônicos da administração pública, garantindo o pleno acesso à informação às pessoas com deficiência (Brasil, 2024, *online*).

Em 2011, o Brasil tornou-se cofundador da Parceria para Governo Aberto (*Open Government Partnership*), que "é uma iniciativa internacional da qual participam diversos países e entidades da sociedade civil, com a proposta de que as entidades governamentais participantes assumam compromissos de transparência em relação aos dados públicos" (Senado Federal, 2024). Após essa parceria, foi promulgada a Lei Nº 12.527, em 18 de novembro de 2011, conhecida como Lei de Acesso à Informação (LAI), com o objetivo de garantir o acesso à informação previsto na Constituição Federal de 1988.

Malin (2013) visualiza os acontecimentos de 2011 como dois marcos no Brasil que afetaram o regime das informações que são criadas e que circulam entre Estado e sociedade: a LAI, com o objetivo de garantir o acesso à informação previsto na Constituição Federal, por meio de iniciativas de transparência ativa e passiva, e a Parceria Governo Aberto (OGP), que acontece a nível internacional, com a participação de diversos países e organizações da sociedade civil, partilhando da ideia de um governo acessível, mais responsivo e mais responsável diante dos cidadãos.

Em abril de 2012, o governo regulamenta a publicação de dados abertos mediante instrução normativa, em que a INDA é definida como a metodologia a ser utilizada pelos órgãos públicos para divulgação de dados abertos no Portal Brasileiro de Dados Abertos (Brasil, 2024, *online*). Em maio do mesmo ano, ocorreu a publicação do documento "Padrão de Dados - Integração de Protocolos do Governo Federal", com o objetivo de estabelecer diretrizes para a

disponibilização de dados em um canal único do governo federal. A consolidação dessas ações refletiu na disponibilização do Portal Brasileiro de Dados Abertos, que foi elaborado de forma colaborativa com a sociedade civil e utilizou de plataformas abertas, como o *software* livre (*Comprehensive Knowledge Archive Network - CKAN)* disponibilizado pela *Open Knowledge Foundation* (OKFN) (Brasil, 2024, *online*).

Ainda em 2012, no mês de junho, houve o Fórum Intergovernamental de Gestores de Geoinformação, no qual foi apresentada a Infraestrutura Nacional de Dados Espaciais (INDE), uma plataforma que visa reunir, catalogar, integrar, harmonizar e disseminar dados e informações geoespaciais do governo brasileiro (Brasil, 2024, *online*).

No ano 2014, foi lançada a versão 1.1 do Padrão de Metadados do Governo Eletrônico (e-PMG), que objetiva garantir que os cidadãos consigam ter acesso rápido e eficiente a descrições de recursos ao pesquisar informações do governo brasileiro na web.

Em 11 de maio de 2016, foi publicado o Decreto Nº 8.777, que instituiu a Política de Dados Abertos do Poder Executivo Federal, que, entre vários objetivos, busca promover a publicação de dados de órgãos e entidades da administração pública federal em formato aberto, fortalecer a cultura de transparência pública e facilitar o intercâmbio de dados entre órgãos e entidades federais e as diferentes esferas da federação (Brasil, 2024, *online*).

Em 25 de novembro de 2022, publica-se o Decreto Nº 11.266, que trata sobre a governança no compartilhamento de dados no âmbito da administração pública federal, e, ao longo dos anos seguintes, foram concretizadas outras iniciativas importantes para o cenário de dados governamentais, como a Estratégia Brasileira para a Transformação Digital 2022-2026 e o Decreto Nº 12.198 de 24, de setembro de 2014, que instituíram a Estratégia Federal de Governo Digital para o período de 2024 a 2027 e a INDA . Os documentos especificamente relacionados a esta dissertação foram mapeadas no segundo objetivo específico e apresentados no Quadro 13, adaptado de Observatório de Dados Abertos (2024), em que foram indicados os elementos relacionados à organização da informação, que foram apresentados nos parágrafos a seguir, conforme categorização: Leis; Decretos; Normativas e Portarias e Instrumentos.

Quadro 13 - Iniciativas de dados abertos no governo federal brasileiro

Iniciativa	Data	Elementos relacionados à organização da informação
Leis		
Lei Nº 12.527/2011 - Lei de Acesso à Informação (LAI)	18 de novembro de 2011	Conceito de dados/informação
Lei Nº 13.709/2018 - Lei Geral de Proteção de Dados Pessoais (LGPD)	14 de agosto de 2018	Definição de dado pessoal Diretrizes para tratamento de dados pessoais Definição de banco de dados
Lei Nº 14.129/2021 - Princípios, regras e instrumentos para o Governo Digital	29 de março de 2021	Garantia de interoperabilidade de sistemas e promoção de dados abertos Definição de dados abertos, dado acessível e formato aberto
Decretos		
Decreto Nº 6.666/2008 - Institui a Infraestrutura Nacional de Dados Espaciais - INDE	27 de novembro de 2008	Definição de dado espacial Conceito de metadados de informações geoespaciais
Decreto Nº 7.724/2012 - Regulamenta a Lei Nº 12.527	16 de maio de 2012	Conceito de dados processados Procedimentos para classificação da informação Conceito de tratamento da informação
Decreto Nº 8.777/2016 - Política de Dados Abertos	11 de maio de 2016	Conceito de dados mais detalhado Conceito de dados abertos Definições sobre Plano de Dados Abertos (PDA)
Decreto Nº 10.046/2019 - Dispõe sobre a governança no compartilhamento de dados no âmbito da administração pública federal e institui o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados	9 de outubro de 2019	Conceito de informações Conceito de interoperabilidade Conceito de governança de dados Regras de compartilhamento de dados
Decreto Nº 11.529/2023 - Institui o Sistema de Integridade, Transparência e Acesso à Informação da Administração Pública Federal e a Política de Transparência e Acesso à Informação da Administração Pública Federal	16 de maio de 2023	Designa ao SITAI a competência de desenvolver e disponibilizar procedimentos, padrões, metodologias e sistemas informatizados que permitam a disseminação de informações públicas
Normativas e Portarias		
Portaria Nº 5/2005 - Institucionaliza os Padrões de Interoperabilidade de Governo Eletrônico	14 de julho de 2005	-

Instrução Normativa Nº 4/2012 - Infraestrutura Nacional de Dados Abertos - INDA	12 de abril de 2012	Política para garantir e facilitar o acesso aos dados e informações produzidas ou custodiadas pelo Poder Executivo Federal
Portaria Nº 92/2014 - Institui a arquitetura e-PING	24 de dezembro de 2014	Definiu um conjunto mínimo de premissas e especificações para utilização de TICs na interoperabilidade de serviços de Governo Eletrônico
Resolução Nº 3/2017 da CGINDA - Aprova as normas sobre elaboração e publicação de Planos de Dados Abertos	13 de outubro de 2017	Estabelece as normas para a elaboração e publicação de Planos de Dados Abertos (PDAs)
Plano de Ação da INDA	Agosto de 2023	Definição de modelo de catalogação de conjuntos de dados
Instrumentos		
Padrão de Metadados do Governo Eletrônico - e-PMG	Última atualização: julho de 2014	Conjunto mínimo de elementos com os dados necessários para recuperação e gerenciamento de informações do governo
Vocabulário de Controlado Governo Eletrônico - VCGE	Última atualização: 2016	Vocabulário controlado criado para indexar informações e ser uma interface de comunicação com o cidadão
Painel de Monitoramento de Dados Abertos do Governo Federal	6 de julho de 2017	Verifica se os órgãos do governo federal têm disponibilizado as informações contidas em suas bases de forma aberta
Padrão de Interoperabilidade de Governo Eletrônico - e-PING	Última atualização: 2018	Incentiva a utilização de vocabulários controlados, taxonomias, ontologias e outros métodos de organização e recuperação de informações
Portal Brasileiro de Dados Abertos	Última atualização: 2022	Instrumento oficial de disponibilização dos DGA do governo federal brasileiro

Fonte: dados da pesquisa e Observatório de Dados Abertos (2024).

Em relação às principais legislações mapeadas está a Lei Nº 12.527/2011, LAI, que incentiva a divulgação de informações de interesse público, independentemente de solicitações (Brasil, 2011), fortalecendo a transparência passiva no país. Em relação à temática da pesquisa, é importante observar as definições adotadas pela lei para alguns conceitos utilizados nesta dissertação. A Lei considera informação como: "dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato" (Brasil, 2011, *online*). Observa-se que os termos informação e dados são considerados sinônimos, contrariando as conceituações em diversas disciplinas, inclusive na CI.

O Decreto Nº 7.724/2012 complementa a LAI com a definição de dados processados como aqueles que são "submetidos a qualquer operação ou tratamento por meio de processamento eletrônico ou por meio automatizado com o emprego de tecnologia da informação" (Brasil, 2012, *online*). Também são estabelecidos procedimentos para a classificação da informação, como a obrigatoriedade de um Termo de Classificação de Informação - TCI (Brasil, 2012, *online*).

Em relação ao tratamento da informação, define-se como: "conjunto de ações referentes à produção, recepção, classificação, utilização, acesso, reprodução, transporte, transmissão, distribuição, arquivamento, armazenamento, eliminação, avaliação, destinação ou controle da informação". Na CI, o tratamento da informação é tido como um processo da organização da informação, que possui duas dimensões, a descritiva e a temática, e compreende "métodos, técnicas e instrumentos de organização do conhecimento e representação de informação, assim como de recuperação e disseminação" (Felix *et al*, 2016, p. 9).

A Lei Nº 13.709/2018 (LGPD) traz definições sobre o tratamento de dados pessoais, aplicáveis à União, Estados, Distrito Federal e Municípios, com o objetivo de proteger os "direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural" (Brasil, 2018, *online*). Define-se dado pessoal como "informação relacionada à pessoa natural identificada ou identificável", e banco de dados como "conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico" (Brasil, 2018, *online*). Destaca-se que essa Lei explicita o tratamento de dados pessoais, elementos não contidos nos dados abertos, mas que podem envolver esse tratamento anterior somado à anonimização ou aos tratamentos que retiram dados sensíveis para se tornarem abertos.

A Lei Nº 14.129/2021 estabelece princípios, regras e instrumentos para aumentar a eficiência da administração pública, principalmente por meio de desburocratização, inovação, transformação digital e participação cidadã (Brasil, 2021, *online*). Existe a previsão de disponibilização de informações e serviços públicos em plataforma única e a garantia de interoperabilidade de sistemas e a promoção de dados abertos (Brasil, 2021, *online*).

Essa lei introduz o conceito de 'base nacional de serviços públicos' como aquela que contém as informações necessárias sobre a oferta de serviços públicos de todos os seus prestadores; de 'plataformas de governo digital' como ferramentas digitais e serviços aos órgãos, geralmente ofertados de forma centralizada e compartilhada, e de 'governo como plataforma' sendo a infraestrutura tecnológica facilitadora do uso de dados de acesso público e promotora da interação entre agentes de forma segura, eficiente e responsável (Brasil, 2021, *online*).

Os conceitos de dados abertos, de dado acessível ao público e de formato aberto, constantes na Lei Nº 14.129/2021, são convergentes com as definições apresentadas na LAI. Essa lei prevê também determinações sobre a abertura de dados e a sua interoperabilidade entre órgãos públicos (Brasil, 2021, *online*).

Em relação aos Decretos, encontra-se o Decreto Nº 6.666/2008, que instituiu a Infraestrutura Nacional de Dados Espaciais - INDE, apresentada como um "conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal" (Brasil, 2008, *online*), e iguala o conceito de dado e informação geoespacial, considerando-o como:

> Aquele que se distingue essencialmente pela componente espacial, que associa a cada entidade ou fenômeno uma localização na Terra, traduzida por sistema geodésico de referência, em dado instante ou período de tempo, podendo ser derivado, entre outras fontes, das tecnologias de levantamento, inclusive as associadas a sistemas globais de posicionamento apoiados por satélites, bem como de mapeamento ou de sensoriamento remoto (Brasil, 2008, *online*).

O Decreto Nº 6.666/2008 também apresenta o conceito de metadados de informações geoespaciais como o "conjunto de informações descritivas sobre os dados, incluindo as características do seu levantamento, produção, qualidade e estrutura de armazenamento,

essenciais para promover a sua documentação, integração e disponibilização, bem como possibilitar a sua busca e exploração" (Brasil, 2008, *online*).

O Decreto Nº 8.777/2016 traz uma definição de dados mais individualizada, considerando dado como uma "sequência de símbolos ou valores, representados em qualquer meio, produzidos como resultado de um processo natural ou artificial" (Brasil, 2016, *online*), e também uma definição que poderia se aplicar ao conceito de DGA, a de dado acessível ao público, descrito como "qualquer dado gerado ou acumulado pelo Governo que não esteja sob sigilo ou sob restrição de acesso nos termos da Lei nº 12.527, de 18 de novembro de 2011" (Brasil, 2016, *online*).

O Decreto Nº 8.777/2016 também traz a definição tradicional de dados abertos, classificando-os como aqueles que são "acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte" (Brasil, 2016, *online*).

Outro conteúdo importante para o contexto da pesquisa, exposto neste decreto, é em relação ao Plano de Dados Abertos (PDA), um documento que orienta as ações de abertura de dados de cada órgão ou entidade da administração pública federal, estabelecendo padrões mínimos de qualidade, visando facilitar o entendimento e a reutilização das informações (Brasil, 2016, *online*).

O Decreto Nº 8.777/2016 estabelece princípios e diretrizes que são consonantes com aqueles relacionados aos dados abertos ao exigir, em seu artigo 3º:

> I - observância da publicidade das bases de dados como preceito geral e do sigilo como exceção;
> II - garantia de acesso irrestrito às bases de dados, as quais devem ser legíveis por máquina e estar disponíveis em formato aberto;
> III - descrição das bases de dados, com informação suficiente para a compreensão de eventuais ressalvas quanto à sua qualidade e integridade;
> IV - permissão irrestrita de reuso das bases de dados publicadas em formato aberto;
> V - completude e interoperabilidade das bases de dados, as quais devem ser disponibilizadas em sua forma primária, com o maior grau de granularidade possível, ou referenciar as bases primárias, quando disponibilizadas de forma agregada;
> VI - atualização periódica, de forma a garantir a perenidade dos dados, a padronização de estruturas de informação e o valor dos dados à sociedade e atender às necessidades de seus usuários; e

> VII - designação clara de responsável pela publicação, atualização, evolução e manutenção de cada base de dado aberta, incluída a prestação de assistência quanto ao uso de dados (Brasil, 2016, *online*).

Apoiando essas ações, em 2017, o Comitê Gestor da Infraestrutura Nacional de Dados Abertos aprovou, por meio da Resolução Nº 3, normas sobre a elaboração e publicação de Planos de Dados Abertos (Brasil, 2017, p. 54).

O Decreto Nº 10.046/2019, que dispõe sobre a governança no compartilhamento de dados na administração pública federal, define informações como "dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato", e a interoperabilidade como "capacidade de diversos sistemas e organizações trabalharem em conjunto, de modo a garantir que pessoas, organizações e sistemas computacionais troquem dados", além de definir a governança de dados como o "exercício de autoridade e controle que permite o gerenciamento de dados sob as perspectivas do compartilhamento, da arquitetura, da segurança, da qualidade, da operação e de outros aspectos tecnológicos" (Brasil, 2019, online).

O Decreto Nº 11.529/2023, que institui o Sistema de Integridade, Transparência e Acesso à Informação da Administração Pública Federal (Sitai) e a Política de Transparência e Acesso à Informação da Administração Pública Federal, tem como um de seus objetivos aumentar a harmonia de informações e dados nas relações entre a administração pública federal e a sociedade, e, para tanto, estabelece algumas competências relacionadas ao Sitai, que são observadas na execução das iniciativas de dados abertos do Poder Executivo Federal como:

> - [...] desenvolver e disponibilizar procedimentos, padrões, metodologias e sistemas informatizados que permitam a disseminação, a obtenção, a utilização e a compreensão de informações públicas;
> - monitorar o atendimento às solicitações de acesso à informação e o cumprimento das obrigações de transparência ativa e de abertura de dados;
> - estimular e apoiar a adoção de medidas de integridade, transparência e acesso à informação para o fortalecimento das políticas públicas;
> - definir critérios e indicadores para a avaliação e o monitoramento da implementação da Política de Transparência e Acesso à Informação da Administração Pública Federal;
> - promover o uso dos dados e das informações públicas pela sociedade para a melhoria da gestão, das políticas e dos serviços; e
> - identificar bases de dados e de informações de interesse público e, conforme o caso, sugerir às unidades setoriais a abertura em transparência ativa (Brasil, 2023, *online*).

A aplicação desse decreto pode ser observada nos PDAs das instituições, no Painel de Monitoramento de Dados Abertos, no Portal Brasileiro de Dados Abertos, no Portal da Transparência, no Plano de Ação da CGU - responsável pela Infraestrutura Nacional de Dados Abertos -, entre outros.

Em relação às Normativas e Portarias, em 2005, foi publicada a Portaria Normativa Nº 5, que institucionalizou os Padrões de Interoperabilidade de Governo Eletrônico (e-PING), que têm o objetivo mais administrativo de estabelecer funções e atribuir responsabilidades (Brasil, 2005, *online*), e foi revogada em 2014, quando foi publicada a Portaria Nº 92, que instituiu a arquitetura e-PING, definindo um conjunto mínimo de premissas, políticas e especificações técnicas, regulamentado para a utilização de Tecnologias de Informação e Comunicação (TICs) na interoperabilidade de serviços de Governo Eletrônico, sendo a sua adoção facultativa e revogando a Portaria Normativa Nº 5 (Brasil, 2014, p. 50).

O Padrão e-PING teve sua última versão disponibilizada em 2018 (https://eping.governoeletronico.gov.br/) e tem algumas políticas gerais norteadoras, como: adoção preferencial de padrão aberto, uso de *software* público ou *software* livre e existência de suporte de mercado (Brasil, 2024, *online*).

A arquitetura e-PING define um conjunto mínimo de premissas, políticas e especificações técnicas para regulamentar a utilização das TICs na interoperabilidade de serviços de Governo Eletrônico e estabelece condições de interação com os demais Poderes (Executivo e Judiciário), as esferas de governo e a sociedade em geral (Brasil, 2018, *online*).

A e-PING aborda interconexão, segurança, meios de acesso, organização e intercâmbio de informações e áreas de integração para Governo Eletrônico. Ela incentiva, também, a utilização de vocabulários controlados, taxonomias, ontologias e outros métodos de organização e recuperação de informações, com o objetivo de facilitar o cruzamento de dados de diferentes fontes de informação (Brasil, 2018, *online*).

A arquitetura recomenda, ainda, a utilização de RDF e RDF *Schema* para a descrição de recursos, a utilização de lista de vocabulários recomendados pelo W3C e pelo governo, o SKOS como sistema de organização de conhecimento e a OWL como linguagem de definição de ontologias (Brasil, 2018, *online*).

Em 2012, foi apresentada a Instrução Normativa Nº 4, que instituiu a já citada INDA, uma política para garantir e facilitar o acesso aos dados e às informações produzidas ou custodiadas

pelo Poder Executivo Federal por parte da sociedade e das instâncias do setor público (Brasil, 2012, p. 1). Nesse documento, são apresentados conceitos que ainda são válidos nas políticas atuais:

> I – dado: sequência de símbolos ou valores, representados em algum meio, produzidos como resultado de um processo natural ou artificial;
> II – informação: conjunto de dados organizados de tal forma que tenham valor ou significado em algum contexto;
> III – dado público: qualquer dado gerado ou sob a guarda governamental que não tenha o seu acesso restrito por legislação específica;
> IV – formato aberto: formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização;
> V – licença aberta: acordo de fornecimento de dados que conceda amplo acesso para que qualquer pessoa os utilize, os reutilize, e os redistribua, estando sujeito a, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença;
> VI – dados abertos: dados públicos representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na rede mundial de computadores e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento; e
> VII – metadado: informação que descreve características de determinado dado, explicando-o em certo contexto de uso (Brasil, 2012, p. 2-3).

Existem também algumas definições que se relacionam diretamente com a representação de dados abertos do governo, como a atribuição ao Comitê Gestor de elaborar o Plano de Ação da INDA, contendo, entre outros, prazos, regras e formas para disponibilização de dados e metadados no Portal Brasileiro de Dados Abertos (Brasil, 2012, p. 5).

Com relação aos planos de ação, o último plano de ação da INDA foi publicado em agosto de 2023, com validade até 2025, e prevê a realização de sete ações para fortalecer a Política de Dados Abertos do Executivo Federal, assim como os demais instrumentos e ferramentas governamentais nessa área (Brasil, 2023). São essas ações:

- *Workshops* de dados governamentais abertos;
- Promoção da inovação com dados abertos;
- Programa continuado de orientação e capacitação em dados abertos;
- Realizar pesquisa sobre reutilização de dados governamentais abertos;
- Revisar *framework* orientativo sobre dados abertos;
- Definir modelo de catalogação de conjuntos de dados (inventário de dados);
- Sistematização de recebimento e implementação de sugestões estruturais para correção, melhorias e validação de dados das bases de dados federais.

Em relação aos instrumentos mapeados, o Padrão de Metadados do Governo Eletrônico (e-PMG) e o Vocabulário Controlado de Governo Eletrônico (VCGE) foram apresentados na subseção 3.3.3, e o Portal Brasileiro de Dados Abertos, na subseção 5.2.2, por estarem intrinsecamente ligados à temática da pesquisa e terem sido explorados mais detalhadamente.

O Painel de Monitoramento de Dados Abertos é uma ferramenta visual criada pela CGU e apresenta um panorama da abertura de dados no Poder Executivo Federal, com atualização mensal (Brasil, s.d.a, *online*). O painel permite o acompanhamento das bases de dados já publicadas, dos cronogramas de publicação e do cumprimento da política de dados abertos por diversos órgãos e entidades do Governo Federal, utilizando como fonte de informação os Planos de Dados Abertos (PDAs) publicados pelas instituições e analisados pela CGU (Brasil, s.d.a, *online*).

Esse painel é um instrumento de planejamento e monitoramento importante, porém não foi objeto da presente pesquisa, pois não apresenta informações sobre os conjuntos de dados publicados, focando no acompanhamento do PDA das instituições do Poder Executivo Federal. A Figura 8 apresenta a visão geral do painel, referente ao mês de outubro de 2024:

Figura 8 - Painel de monitoramento de dados abertos

Fonte: Brasil (s.d.a).

Na próxima subseção, apresenta-se o instrumento oficial de disponibilização dos conjuntos de dados e recursos do Poder Executivo Federal: o Portal Brasileiro de Dados Abertos, que foi utilizado nesta pesquisa.

5.2.2 Portal Brasileiro de Dados Abertos

Nesta seção, apresenta-se uma caracterização do Portal Brasileiro de Dados Abertos, atendendo ao terceiro objetivo específico desta dissertação, que é uma ferramenta oficial de disponibilização de dados e informações públicas do Governo Federal. Ele se propõe a ser um serviço simplificado que organiza e padroniza o acesso aos dados públicos, priorizando o seu reúso e a utilização de tecnologias modernas (Brasil, s.d.f, *online*).

Segundo a seção de perguntas frequentes do Portal Brasileiro de Dados Abertos (Brasil, s.d.f, *online*), os objetivos do portal são: promover a interlocução entre atores da sociedade e o governo, pensando na melhor utilização dos dados, resultando em impactos positivos sob os pontos de vista social e econômico; disponibilizar dados relativos a diversas temáticas da administração pública e ser ponto referencial e único para a busca e o acesso a dados públicos brasileiros de todo e qualquer assunto ou categoria.

O portal funciona como um catálogo de dados federado, facilitando a busca e utilização de dados publicados pelos órgãos do governo, que os disponibilizam conforme cronogramas estabelecidos em seus Planos de Dados abertos (PDAs) (Brasil, s.d.f, *online*). Tanto a elaboração do PDA quanto a publicação dos dados é responsabilidade de cada órgão.

A partir do Decreto Nº 8.777/2016 e do Decreto Nº 9.903/2019, foi designada à CGU a atribuição de monitoramento do cumprimento das diretrizes e obrigações das normativas por parte dos órgãos do Poder Executivo Federal, além da gestão do portal (Brasil, s.d.f, *online*).

Na data de execução da pesquisa (28/07/2024), constavam no portal 13.231 conjuntos de dados, 83.694 recursos, 10 reúsos, 263 organizações e 51.606 usuários como demonstrado na Figura 9:

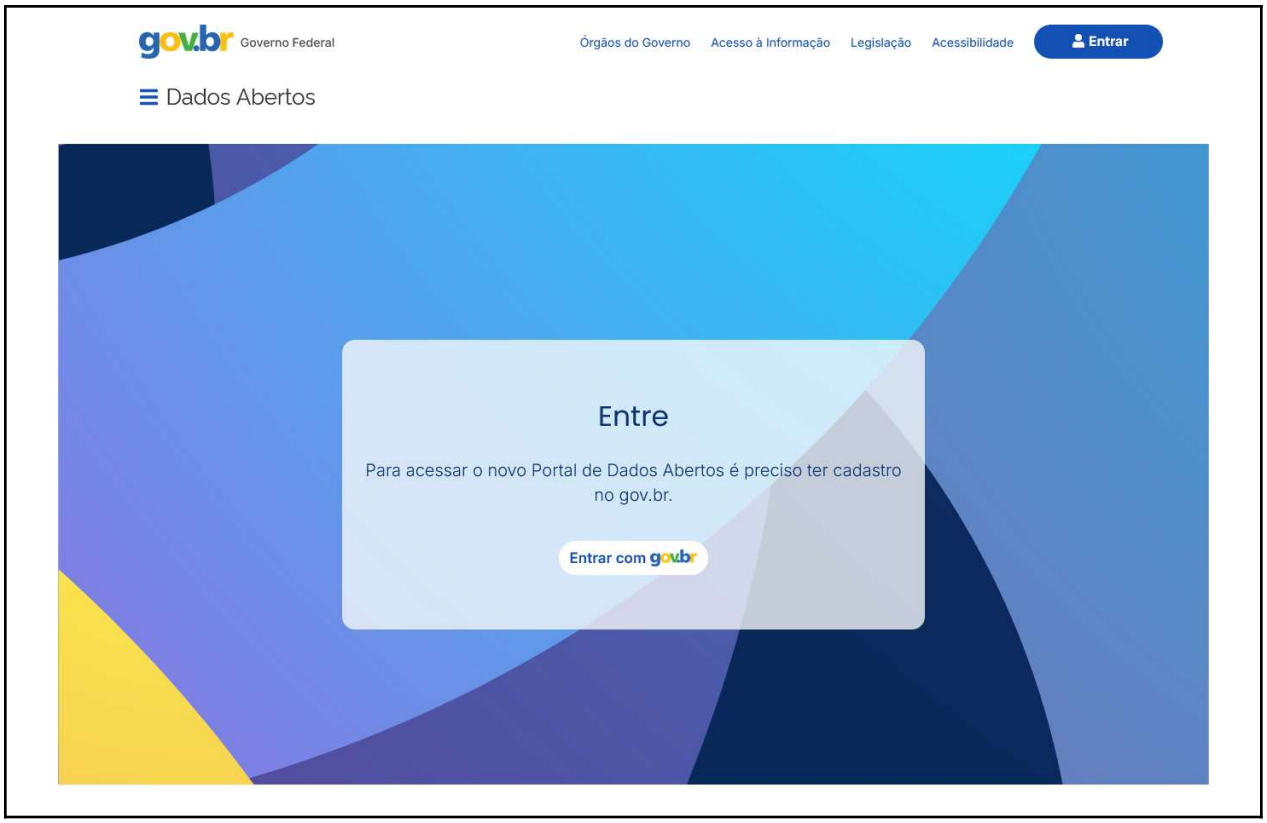
Figura 9 - Página inicial do Portal Brasileiro de Dados Abertos

Fonte: Brasil (s.d.c, *online*).

Chama a atenção o fato de o portal exigir identificação (login e senha) do usuário para realizar algumas ações, como demonstrado na Figura 10, o que poderia gerar questionamentos acerca dos princípios dos dados abertos. Conforme apontado por Nascimento Silva (2024), novos recursos e funcionalidades disponibilizados na última versão do Portal, em novembro de 2022, permitem a interação do usuário com os conjuntos de dados, como: seguir, avaliar, realizar contato, sugerir *tag*, reportar problema, sugerir melhoria, incluir um tópico de discussão, mas, ao acessá-los, é exigida a autenticação com uma conta GOV BR.

Nascimento Silva (2022) observou a exigência de autenticação do usuário em outros instrumentos da Política de Dados Abertos do Brasil, estando em oposição ao sexto princípio dos DGA. Conforme os princípios criados em 2007, o acesso aos dados deve ser não discriminatório, dessa forma, os dados devem estar disponíveis para todos, sem exigência de requerimento ou cadastro. Uma exceção seria válida para o controle de acesso dos órgãos publicadores de dados, mas para funcionalidades que o cidadão acessa não foi encontrada uma justificativa no Portal.

Figura 10 - Solicitação de autenticação no Portal



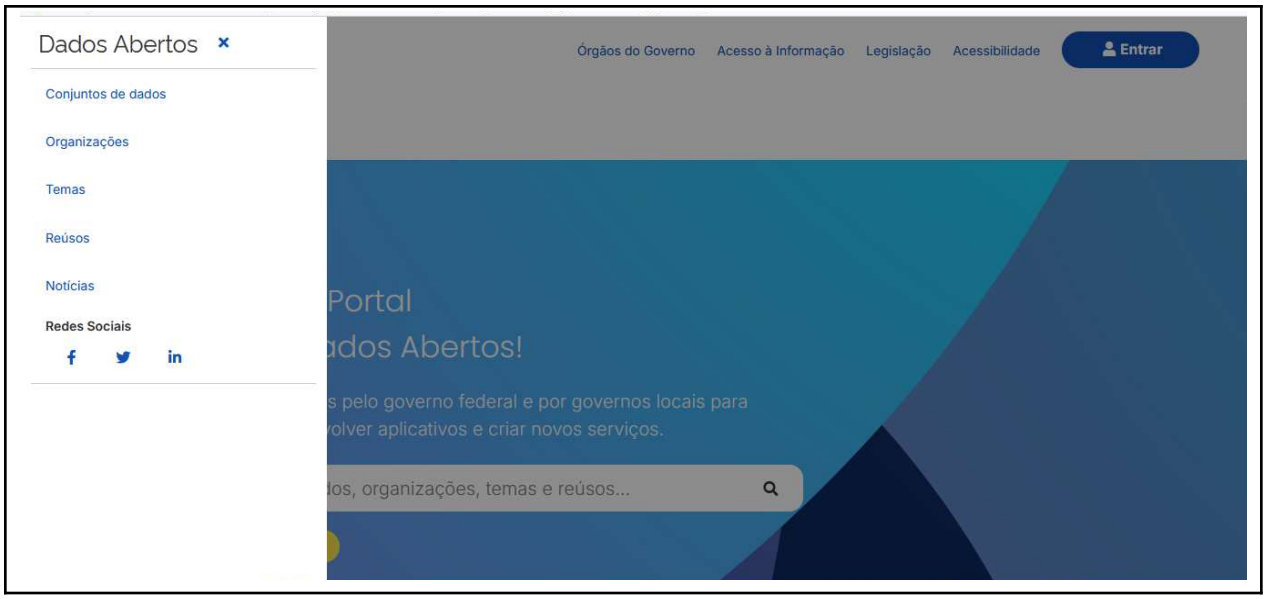
Fonte: Brasil (s.d.c).

Com relação à categorização existente, o portal apresenta, em sua página inicial, uma caixa de busca com as opções: conjuntos de dados, organizações, temas e reúsos, como apresentado na Figura 11. As mesmas opções de busca são apresentadas separadamente acessando via menu lateral, como apresentado na Figura 12.

Figura 11 - Opções de busca no Portal

Fonte: Brasil (s.d.c, *online*).

Figura 12 - Menu lateral de navegação no Portal

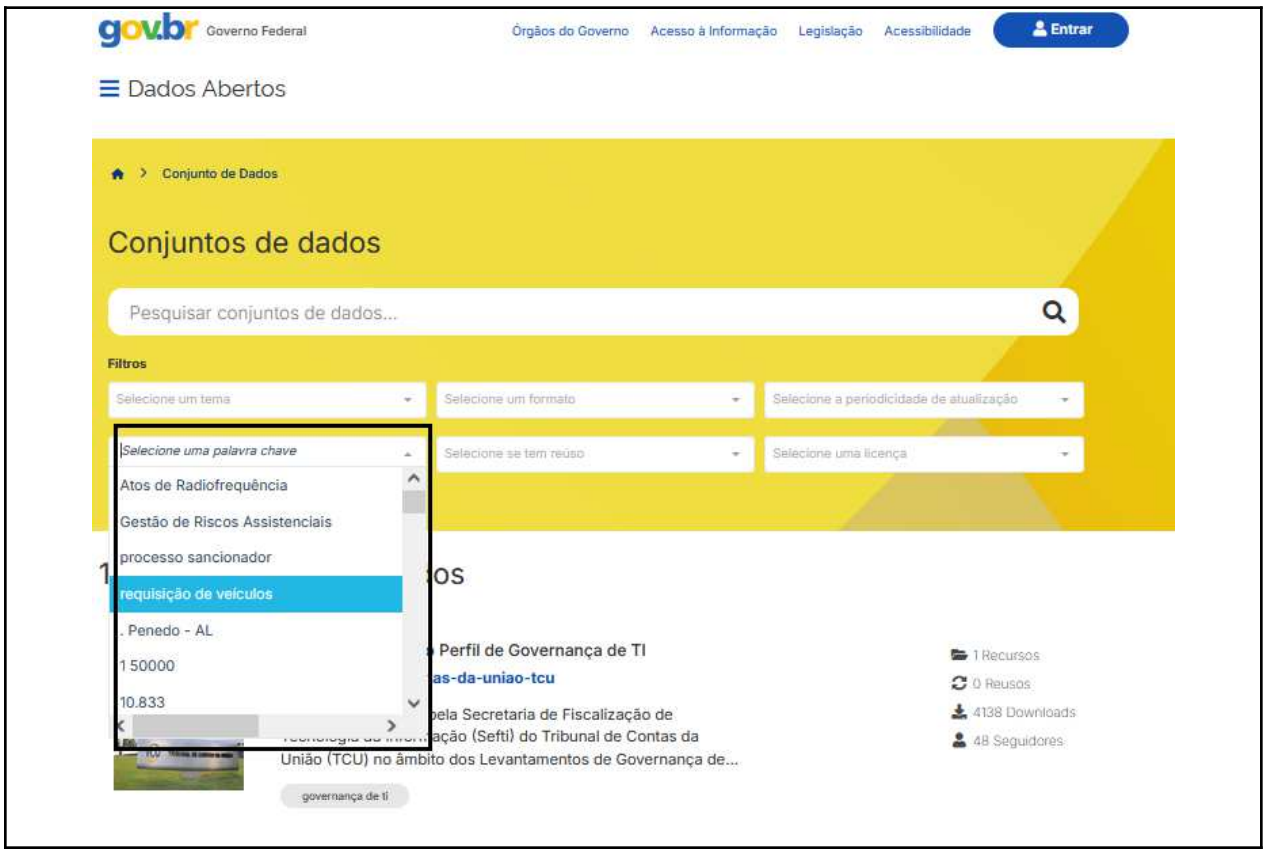


Fonte: Brasil (s.d.c, *online*).

Com relação à recuperação de dados, há uma funcionalidade de pesquisa na opção ‘conjuntos de dados’, que permite o refinamento com a utilização de vários filtros, podendo realizar a filtragem por: temas, formatos, periodicidade de atualização, palavra-chave, reúso, licenças e refinar a busca por organizações, filtrando por esfera, UF ou município, ou pelo nome

da organização. No entanto, ao navegar pelos filtros, observa-se que vários deles apresentam erros no cadastro das opções, como opções repetidas ou com erros ortográficos, que tornam a busca imprecisa, como demonstrado no filtro ‘palavra-chave’, na Figura 13:

Figura 13 - Filtros de busca no Portal



Fonte: Brasil (s.d.c, *online,* grifo nosso).

Em relação a busca por temas, ela apresenta uma categorização com 26 temas, sendo eles: abastecimento, administração, agropecuária, pesca e extrativismo, comunicações, comércio e serviços, cultura, dados estratégicos, defesa nacional, economia e finanças, educação, energia, esporte e lazer, habitação, indústria, infraestrutura e fomento, meio ambiente, pesquisa e desenvolvimento, planejamento e gestão, previdência social, proteção social, relações internacionais, saneamento, saúde, segurança e ordem pública, trabalho, transportes e urbanismo. Esses temas foram definidos de acordo com o VCGE (Versão 2.1.0 - 2016) (Brasil, 2022, *online*).

Os dados catalogados no portal são organizados em conjuntos de dados e recursos, sendo esses conjuntos os elementos principais, retornados a partir das buscas, e cada recurso

assimilando uma fonte de dados (Brasil, s.d.f, *online*). Cada recurso poderia ser um conjunto de dados separado, porém prefere-se que os recursos sejam agrupados quando forem advindos da mesma base de dados ou possuírem metadados em comum, para facilitar a busca e a compreensão do seu conteúdo (Brasil, s.d.f, *online*).

O portal não armazena diretamente os dados que disponibiliza, mas, os metadados associados a eles e os *links* de redirecionamento para os dados, sendo cada órgão responsável pela sua hospedagem e publicação na internet (Brasil, s.d.f, *online*). Sendo assim, o portal cataloga os dados, funcionando como um centralizador que facilita a sua recuperação pelos cidadãos. Para realizar essa função, o portal utiliza o CKAN (Brasil, s.d.f, *online*), que é uma ferramenta livre para a criação de repositórios de dados abertos, funcionando como um sistema de gerenciamento de dados que alimenta *hubs* e portais (Brasil, 2021, *online*).

O portal apresenta um manual para a publicação de dados abertos, disponível no endereço eletrônico: https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-conjunto-de-dados, que contém um dicionário de dados, orientando sobre o preenchimento de cada metadado, sendo obrigatórios os metadados para cadastro do conjunto de dados: título, organização à qual o recurso está associado, descrição, licença de uso, área técnica responsável pelo dado, e-mail da área técnica, temas aos quais o dado está associado, palavras-chave e versão; e obrigatórios em relação ao cadastro de recurso: título, url, tipo de recurso, formato e descrição.

Existe um manual semelhante para o cadastro de reúso, disponível no endereço eletrônico: https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-reuso e que estabelece como obrigatórios os seguintes metadados: nome, URL, descrição, tipo, nome do responsável, organização a qual o reúso está associado, conjuntos de dados de origem e assuntos / palavras-chave.

No Quadro 14, são apresentados o quantitativo de conjuntos de dados e de reúso, organizados por tema, até 30 de agosto de 2024. Observa-se uma inconsistência nos números apresentados no portal, o que prejudica a confiabilidade da coleta de dados, como, por exemplo, na quantidade de organizações, sinalizadas em um total de 263, mas na somatória das organizações, em cada temática, tem-se o número 73.

Quadro 14 - Quantidades de Organizações, Conjuntos de dados e Reúsos no Portal

Tema	Organizações	Conjuntos de dados	Reúsos
Portal de Dados Abertos	263	13.347	10
Abastecimento	0	0	0
Administração	5	161	28
Agropecuária, Pesca e Extrativismo	2	9	2
Comunicações	6	186	2
Comércio e Serviços	5	24	0
Cultura	1	67	0
Dados Estratégicos	3	128	1
Defesa Nacional	6	24	0
Economia e Finanças	4	144	0
Educação	5	603	0
Energia	0	0	0
Esporte e Lazer	0	0	0
Habitação	4	8	0
Indústria	2	5	0
Infraestrutura e Fomento	1	1	0
Meio Ambiente	5	20	0
Pesquisa e Desenvolvimento	2	15	0
Planejamento e Gestão	4	5	0
Previdência Social	1	1	0
Proteção Social	2	44	36
Relações Internacionais	7	10	0
Saneamento	0	0	0
Saúde	1	573	0
Segurança e Ordem Pública	1	15	0
Trabalho	5	80	0
Transportes	1	2	0
Urbanismo	0	0	0

Fonte: elaborado pela autora (2024).

Atualmente, existem documentos como o Plano de Ação da INDA 2023-2025, que sugerem que o Portal passará por atualizações, visando aprimorar a disponibilização dos DGA brasileiros e implementando um modelo de catalogação.

Perspectivas futuras para o Portal incluem a implementação de um modelo de catalogação de conjuntos de dados, criando o ‘Catálogo de Dados’ para suprir a necessidade de fornecer uma solução única de disponibilização de informações sobre bases de dados disponíveis nos órgãos e entidades federais. Conforme o Plano de Dados Abertos da CGU (CGU, 2024), para atingir os objetivos propostos, o Catálogo de Dados adotará o padrão DCAT-BR (do inglês *Data Catalog Vocabulary*), adaptado pela Secretaria de Governo Digital (SGD). Destaca-se que as mudanças sinalizadas ainda estão no nível de planejamento e sua execução ainda não foi implementada. Desta forma, considerou-se, para esta dissertação, a versão vigente até novembro de 2024 do Portal Brasileiro de Dados Abertos. A próxima subseção aborda algumas informações disponibilizadas pelos órgãos responsáveis.

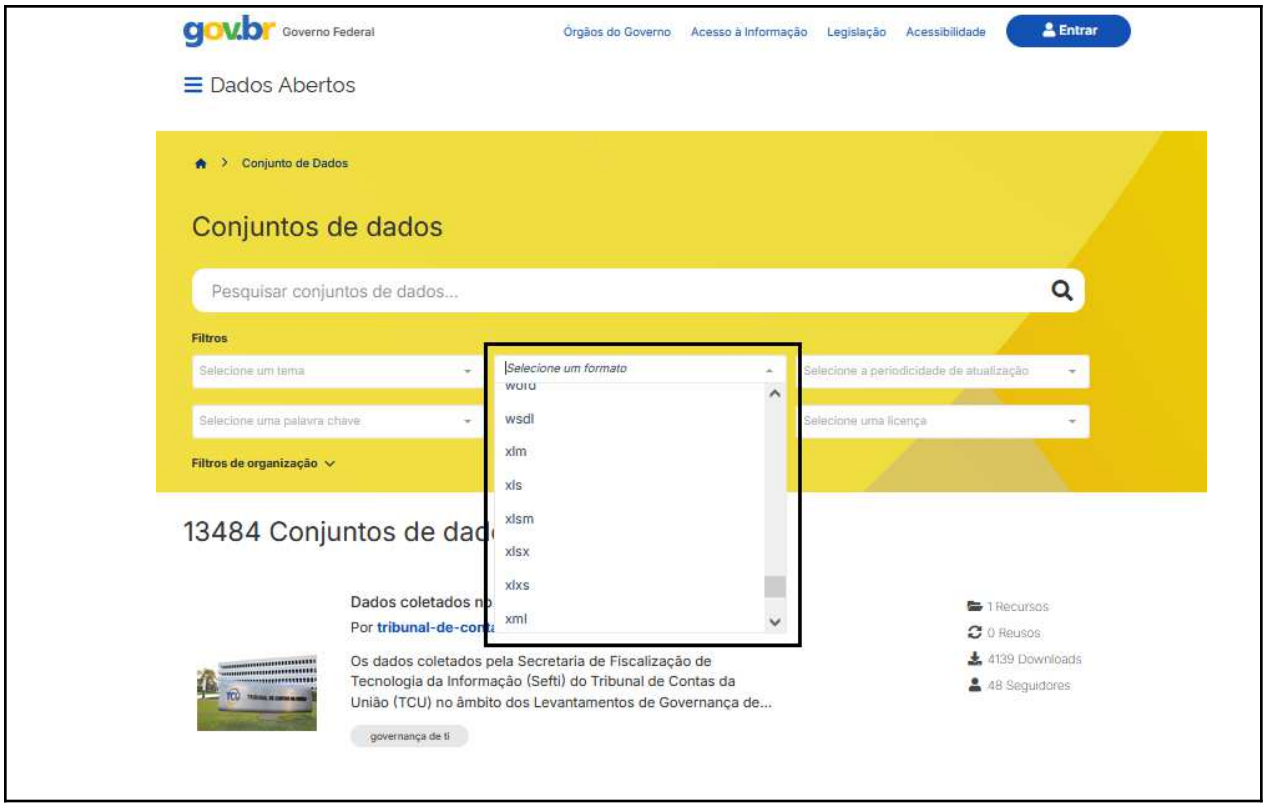
5.2.3 Tipologias de dados abertos, formatos, padrões e metadados utilizados

Em continuidade ao cumprimento do terceiro objetivo específico, apresenta-se nesta subseção a identificação e análise de tipologias de dados abertos, formatos, padrões e metadados utilizados.

O Portal Brasileiro de Dados Abertos disponibiliza a Cartilha Técnica para Publicação de Dados Abertos no Brasil, que orienta sobre as boas práticas para a publicação de dados na internet. Nesse documento não há especificação sobre **tipologias** de dados presentes no Portal. Segundo a Cartilha, a filosofia de dados abertos não apresenta uma lista exaustiva de **formatos** permitidos, porém eles devem ter especificação aberta, não proprietária e serem estruturados. A cartilha lista alguns exemplos que atendem a esses requisitos: JSON, CSV, ODS, RDF, SVG e GML.

Ao realizar observação direta no portal, por meio da utilização dos filtros de busca, percebeu-se que não existe controle de vocabulário no campo de preenchimento do formato do recurso, pois foram identificados nomes de formatos inexistentes, que provavelmente são erros de digitação, como mostrado na Figura 14:

Figura 14 - Filtro de busca por formato do recurso no Portal



Fonte: Brasil (s.d.c, *online,* grifo nosso).

A utilização de **padrão de metadados**, **vocabulários controlados** ou similares é tímida. Não foi localizada, na data de pesquisa, 3 de outubro de 2024, menção ao e-PMG 1.1. A partir da observação direta, infere-se que o VCGE é utilizado somente para controlar os 26 temas de classificação dos conjuntos de dados. Segundo o Portal, o padrão de metadados utilizado é estabelecido na Cartilha Técnica para Publicação de Dados Abertos.

A Cartilha Técnica para Publicação de Dados Abertos no Brasil define um conjunto de metadados obrigatórios e desejáveis a serem utilizados, como apresentado no quadro 15:

Quadro 15 - Metadados da Cartilha Técnica para Publicação de Dados Abertos no Brasil

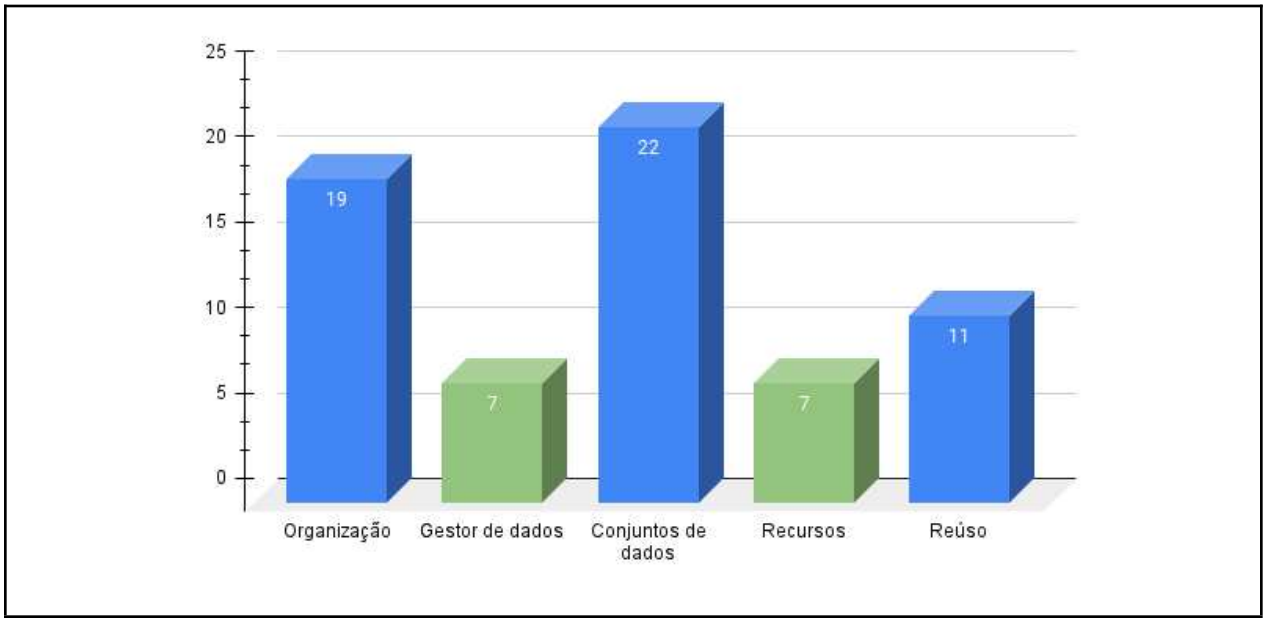
ID	Campo	Descrição	Obrigatório / Desejável
1	Título	Nome do conjunto de dados	Obrigatório
2	Descrição	Breve explicação sobre os dados	Obrigatório
3	Catálogo de origem	Página (URL) do órgão onde está publicado o conjunto de dados	Obrigatório
4	Órgão responsável	**Nome** e **sigla** do órgão ou entidade responsável pela publicação do conjunto de dados	Obrigatório
5	Categoria VCGE	Lista hierarquizada de assuntos do governo que utiliza termos comuns	Obrigatório
6	Identificador (do recurso)	URL persistente que aponta para o recurso na Web	Obrigatório
7	Título (do recurso)	Nome do recurso	Obrigatório
8	Formato (do recurso)	Formato do recurso. Ex.: XML, JSON, CSV, etc	Obrigatório
9	Descrição (do recurso)	Breve detalhamento sobre o conteúdo do recurso	Obrigatório
10	Etiquetas	Lista de palavras-chaves relacionadas ao conjunto de dados	Desejável
11	Autoria	Instituição ou pessoa responsável pela produção do recurso	Desejável
12	Documentação	URL de documento que expõe detalhes sobre o conjunto de dados	Desejável
13	Cobertura geográfica	Localização ou região geográfica a que se referem os dados	Desejável
14	Cobertura temporal	Data ou período a que referem os dados	Desejável
15	Granularidade geográfica	Precisão geográfica da cobertura geográfica	Desejável
16	Granularidade temporal	Precisão temporal da cobertura temporal	Desejável
17	Frequência de atualização	Frequência temporal com que o conjunto de dados é atualizado	Desejável
18	Referências	Relações com outros conjuntos de dados	Desejável
19	Metodologia	Processo de criação dos dados	Desejável
20	Vocabulário / Ontologia	Documentos estruturados com metadados específicos do conjunto de dados	Desejável

Fonte: adaptado de Brasil (2016, p. 7-8).

Dos metadados indicados na Cartilha, todos os obrigatórios foram identificados no Portal, e, quanto aos desejáveis, não foram identificados os campos Documentação, Granularidade temporal, Metodologia e Vocabulário / Ontologia ou similares. Essa análise foi realizada comparando os dados da Cartilha com os manuais para a publicação de conjuntos de dados e reúso disponíveis no Portal.

Foram identificados, em observação direta durante o mês de outubro de 2024, um total de 66 metadados em utilização no Portal, sendo 46 de preenchimento obrigatório, 16 opcionais e quatro de geração automática. Entre eles, 19 são metadados associados ao cadastro da Organização, sete relacionados ao Gestor de Dados, 22 relacionados aos Conjuntos de Dados, sete relacionados aos Recursos e 11 relacionados a Reúso. Essas categorias não são mencionadas nos documentos analisados e foram detectadas por observação direta. A distribuição dos metadados em cada categoria é apresentada no Gráfico 2:

Gráfico 2 - Metadados em utilização no Portal



Fonte: elaborado pela autora (2024).

Mediante a observação direta de uma amostra de conjuntos de dados, incluindo órgãos como: Tribunal de Contas da União, Caixa Econômica Federal, Empresa Brasileira de Pesquisa Agropecuária Brasileira, Força Aérea, Estado de São Paulo - SP, Universidade Federal do Rio Grande do Sul, Instituto Brasileiro do Turismo - EMBRATUR e Fundação Nacional de Saúde,

representando temáticas diversas, páginas das organizações e recursos, detectou-se que 34 metadados ficam visíveis para o usuário do Portal, e para três deles não foi possível inferir se são visíveis ao usuário ou não (Descontinuado, Data da descontinuação e Conjuntos de dados associados). Nem todos os metadados têm rótulos de identificação, como mostrado na Figura 15:

Figura 15 - Identificação de metadados em um conjunto de dados no Portal

Fonte: Brasil (s.d.c, *online,* grifo nosso).

No quadro 16, são apresentados os metadados identificados no Portal, e foi feita uma proposta de sua correspondência com o e-PMG 1.1:

Quadro 16 - Metadados do Portal Brasileiro de Dados Abertos

ID	Campo	Descrição	Nível	Obrigatoriedade (Obrigatório / Opcional)	Visibilidade (Sim/Não)	Correspondência e-PMG 1.1
1	Tipo de organização	Governamental ou não governamental	Organização	Obrigatório	Não	Contexto Jurídico-administrativo (*mandate*)
2	Esfera	Federal, estadual ou municipal		Obrigatório	Sim	Contexto Jurídico-administrativo (*mandate*)
3	Órgão SIORG	Cadastro oficial dos órgãos e entidades da administração direta, autárquica e fundacional do Poder Executivo Federal junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG		Obrigatório	Não	Contexto Jurídico-administrativo (*mandate*)
4	Código SIORG	Referencial numérico que identifica o órgão ou a entidade junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG		Obrigatório	Não	Contexto Jurídico-administrativo (*mandate*)
5	Poder	Executivo, Legislativo, Judiciário ou órgãos essenciais à justiça, como Ministérios Públicos, Defensorias Públicas e Procuradorias		Obrigatório	Sim	Contexto Jurídico-administrativo (*mandate*)
6	Organização Pai	Organização à qual o órgão/entidade esteja vinculado		Obrigatório	Sim	Contexto Jurídico-administrativo (*mandate*)
7	Nome do órgão	Nome do órgão/entidade por extenso		Obrigatório	Sim	Título (*Title*)
8	Imagem	Logomarca oficial do órgão/entidade		Opcional	Sim	-
9	Sigla	Abreviação que identifica o órgão		Obrigatório	Sim	Contexto Jurídico-administrativo (*mandate*)
10	CNPJ	Número designado pela Receita Federal para identificar o órgão/organização		Obrigatório	Sim	Identificador (*Identifier*)
11	UF	Unidade federativa em que o órgão/entidade esteja localizado		Obrigatório	Sim	-

12	Município	Município em que o órgão/entidade esteja localizado		Obrigatório	Sim	-
13	Site da organização	Sítio eletrônico oficial do órgão/entidade		Obrigatório	Não	Contexto Jurídico-administrativo (*mandate*)
14	E-mail da organização	E-mail institucional do órgão/entidade		Obrigatório	Não	Contexto Jurídico-administrativo (*mandate*)
15	Nome da autoridade	Nome da autoridade responsável pelo órgão/entidade		Obrigatório	Não	Título (*Title*)
16	E-mail da autoridade	E-mail institucional da autoridade responsável pelo órgão/entidade		Obrigatório	Não	-
17	Telefone da autoridade	Telefone da autoridade responsável pelo órgão/entidade		Obrigatório	Não	-
18	Status da organização	Ativo ou extinto		Obrigatório	Sim	Contexto Jurídico-administrativo (*mandate*)
19	Descrição (da organização)	Descrição do órgão/entidade, como área de atuação, missão e demais informações que o órgão/entidade achar pertinente compartilhar		Obrigatório	Sim	description.resumo (*description.abstract*)
20	Nome do gestor	Nome do gestor da organização	Gestor de dados	Obrigatório	Não	Título (*Title*)
21	CPF do gestor	CPF do gestor da organização		Obrigatório	Não	Identificador (*Identifier*)
22	Telefone do gestor	Telefone do gestor da organização		Obrigatório	Não	-
23	Imagem do gestor	Imagem do gestor da organização		Opcional	Não	-
24	Data de cadastro do gestor	Data de cadastro do gestor da organização		Automático	Não	data.publicação (*date.issued*)
25	Tipo de perfil	Tipo de perfil de usuário		Automático	Não	Tipo (*Type*)
26	Descrição do perfil	Descrição do perfil de usuário		Opcional	Não	description.resumo (*description.abstract*)
27	Conjunto de dados template	Campos de um conjunto de dados já existente para facilitar a criação de um novo conjunto de dados com características semelhantes	Conjunto de dados	Opcional	Não	-
28	Título do conjunto de dados	Nome do conjunto de dados. A nomenclatura deve ser idêntica à que foi elencada no Plano de Dados Abertos –		Obrigatório	Sim	Título (*Title*)

		PDA, quando esse instrumento for aplicável				
29	Organização à qual o recurso está associado	Campo que identifica a organização que está cadastrando o conjunto de dados		Obrigatório	Sim	criador.autor (*creator.autor*)
30	Item de inventário de origem	Indicar a qual item do inventário o conjunto de dados está relacionado		Opcional	Não	Fonte (*Source*)
31	Descrição do conjunto de dados	Descrição sobre o conjunto de dados, incluindo metadados que o identifique		Obrigatório	Sim	description.resumo (*description.abstract*)
32	Licença de uso	Selecionar sob qual espécie de licença o conjunto de dados será publicado		Obrigatório	Sim	Direitos (*Rights*)
33	Área técnica responsável pelo dado	Indicar qual área técnica do órgão/entidade é responsável pelo conjunto de dados		Obrigatório	Não	criador.produtor (*creator.productor*)
34	E-mail da área técnica responsável	Indicar e-mail de contato da área técnica responsável pelo conjunto de dados		Obrigatório	Não	-
35	Periodicidade de atualização	Indicar qual será a periodicidade com que o conjunto de dados será atualizado		Obrigatório	Sim	data.frequenciaDeAtualização
36	Temas aos quais o dado está associado	Escolha a área temática à qual o conjunto de dados esteja vinculado (VCGE)		Obrigatório	Sim	Assunto.categoria (*subject.category*)
37	Assuntos / Palavras-chave	Resumo dos principais aspectos do conjunto de dados		Obrigatório	Sim	Assunto.palavra-chave (*subject.keyword*)
38	Cobertura temporal início	Data ou período de início dos dados disponibilizados		Opcional	Não	Abrangência temporal
39	Cobertura temporal fim	Data ou período de fim dos dados disponibilizados		Opcional	Não	Abrangência temporal
40	Cobertura espacial	Esfera de cobertura espacial do conjunto de dados		Opcional	Não	Abrangência espacial
41	Granularidade espacial	Nível de granularidade espacial do conjunto de dados		Opcional	Não	Abrangência espacial
42	Versão	Versão do conjunto de dados que está sendo cadastrada		Obrigatório	Não	Identificador (*Identifier*) identificador.versão
43	Atualização de versão	Indicar se houve atualização da versão do conjunto de dados disponibilizado na última alteração realizada		Opcional	Sim	-
44	Visibilidade	Pública ou privada		Obrigatório	Não	-
45	Homologação	Indicar se o conjunto de dados foi homologado pela autoridade competente		Opcional	Não	-

46	Descontinuado	Indicar se conjunto de dados foi descontinuado e não passará por futuras atualizações		Opcional	-	-
47	Data descontinuação	Data a partir da qual o conjunto de dados foi descontinuado		Opcional	-	data.disponibilidade (*date.available*) data.validade (*date.valid*)
48	Conjuntos de dados associados (interoperabilidade)	Indicar outros conjuntos de dados que estejam associados ao conjunto de dados cadastrado		Opcional	-	Relação (*Relation*)
49	Título de recurso	Nome do recurso	Recurso	Obrigatório	Sim	Título (*Title*)
50	URL do recurso	Endereço da web em que o recurso está hospedado		Obrigatório	Sim	-
51	Tipo do recurso	Selecionar se o recurso se refere a dados, documentação, dicionário de dados, API ou outros		Obrigatório	Sim	Tipo de Recurso (*Type*)
52	Formato do recurso	Selecionar sob qual formato o recurso será disponibilizado		Obrigatório	Sim	Formato (*Format*)
53	Descrição do recurso	Breve descrição sobre o recurso		Obrigatório	Sim	descrição.resumo (*description.abstract*)
54	Data de criação do recurso	Data de criação do recurso		Automático	Sim	data.criação (*date.created*)
55	Última alteração do recurso	Última alteração do recurso		Automático	Sim	data.alteração (*date.modified*)
56	Nome do reúso	Nome do reúso por extenso	Reúso	Obrigatório	Sim	Título (*Title*)
57	URL do reúso	Endereço da web onde o reúso está hospedado		Obrigatório	Sim	-
58	Descrição do reúso	Explicação sobre o reúso		Obrigatório	Sim	descrição.resumo *(description.abstract*)
59	Tipo do reúso	Selecionar, entre as opções fornecidas – aplicativo, site, dados e outros –, em qual espécie se encaixa o reúso cadastrado		Obrigatório	Sim	Tipo de Recurso (*Type*)
60	Nome do responsável	Identificação do responsável pelo reúso cadastrado		Obrigatório	Não	criador.produtor (*creator.productor*)
61	Imagem do reúso	Imagem referente ao reúso		Opcional	Sim	-
62	Organização a qual o reúso está associado	Selecionar entre as opções fornecidas, qual órgão/entidade a que o reúso está relacionado		Obrigatório	Sim	autor.criador (*autor.creator*) autor.produtor (*autor.productor*)

63	Conjuntos de dados de origem	Indicar o(s) conjunto(s) de dados que foram utilizados para o reúso		Obrigatório	Sim	Fonte (*Source*)
64	Assuntos / Palavras-chave	Palavras que destacam o tema do reúso		Obrigatório	Sim	Assunto.palavra-chave (*subject.keyword*)
65	Visibilidade	Pública ou privada		Opcional	Não	-
66	Homologação	Indicador se o reúso foi homologado pela autoridade competente		Opcional	Não	-

Fonte: elaborado pela autora (2024).

Foi observado que nem sempre ocorrem correspondências exatas dos metadados identificados, sendo um dos prováveis motivos que justificam o foco do e-PMG 1.1 em fornecer metadados para a descrição de recursos informacionais, e não de um catálogo de dados completo, e os metadados identificados no Portal extrapolam esse escopo, descrevendo também organizações, pessoas, conjuntos de dados e reúsos, além de recursos. Ainda assim, acredita-se que a correlação sugerida no Quadro 16 seria funcional caso fosse colocada em prática, principalmente pelo fato de o e-PMG apresentar uma correlação com diversos outros padrões de metadados, como: e-GMS (*Government Metadata Standard*), AGLS (*Australian Government Locator Service*), GILS (*Government Locator Service*), FGDC (*Federal Geographic Data Committee Metadata*), MARC (*Machine Readable Cataloging*) e MIP (Metainformação para Interoperabilidade).

Percebeu-se que o governo federal tem atuado em diversas frentes para garantir que os princípios de dados abertos sejam cumpridos e que ferramentas para disponibilização dos dados sejam implementadas e fornecidas para utilização do público. Porém, é muito importante que, além do engajamento no desenvolvimento de novas ferramentas e aplicações, sejam avaliadas a usabilidade, a validade, a pertinência e a adequação do que já existe, inclusive para evitar retrabalho e incentivar o reúso. Por isso, na próxima seção, apresenta-se uma proposta de metadados que leva em consideração todo o contexto existente, além de recomendações e boas práticas mapeadas na literatura científica.

6 PROPOSTA DE MODELO DE METADADOS - METAOPENGOV

Nesta seção, apresenta-se a consolidação da proposta para cumprimento do quarto objetivo específico da pesquisa: "**Propor uma estrutura de representação descritiva para os dados abertos do governo federal brasileiro no Portal Brasileiro de Dados Abertos com a utilização de metadados"**. A partir de aprendizados adquiridos na construção da fundamentação teórica e na sistematização das etapas 1, 2 e 3, foram formuladas recomendações que contemplam a variedade e as especificidades dos dados abertos do governo federal brasileiro, a fim de aprimorar a representação desses e, consequentemente, sua recuperação.

Para a apresentação dessa proposta, foram considerados três pilares, estabelecidos durantes a execução das etapas anteriores:

- Legislação e documentação (Etapa 2):
    - Legislações, normativas, portarias e diretrizes sobre acesso à informação;
    - Política de Dados Abertos e legislações complementares.
- Padrões e instrumentos de organização e representação da informação (Fundamentação teórica e Etapa 2):
    - Padrões de metadados (e-PMG, DC e DCAT-3);
    - Vocabulários (VCGE);
    - Boas Práticas para Dados na Web (Lóscio; Burle; Calegari, 2019).
- Estrutura existente no Portal (Etapa 3):
    - Aspectos, funcionalidades e metadados da estrutura existente.

Inicialmente, por meio da identificação da legislação e documentação, foi possível traçar um panorama das diretrizes existentes para conduzir a gestão de dados abertos no governo federal brasileiro e evidenciar as informações básicas a serem disponibilizadas obrigatoriamente, de acordo com a legislação. A seguir, foram estudados os instrumentos existentes que poderiam trazer contribuições para a otimização da representação descritiva do Portal, e, por fim, analisou-se a estrutura atual, na perspectiva de usuária comum, para identificar aspectos passíveis de melhorias. Nas subseções seguintes, será detalhado o processo que culminou na proposta de metadados apresentada.

6.1 Detalhamento da proposta

Nesta seção, serão apresentados os elementos considerados na proposta de metadados, que envolveram aspectos sobre: estrutura e organização do Portal, metadados sugeridos de acordo com níveis de descrição propostos e sua correspondência com o padrão DCAT-3.

6.1.1 Estrutura e organização

A partir do mapeamento realizado no Portal, na subseção 5.2.2, foram identificados elementos da estrutura de organização atual para levantar lacunas e realizar proposições de melhoria. As recomendações desta pesquisa visam ao aprimoramento da representação dos DGA do governo federal, de forma a impactar positivamente o **usuário consumidor**, ou seja, o cidadão, por isso, trabalhar-se-á com os itens visíveis para esse usuário.

Foi identificado, na estrutura atual do Portal, que existe uma separação em blocos de preenchimento para as informações. Essa estrutura pode ser observada nos manuais para publicadores de dados. Dessa forma, a proposta sugere o **agrupamento por categorias**, em que os metadados relacionados possuem características comuns, tendo sido identificadas as seguintes categorias: Organização, Gestor de Dados, Conjunto de Dados, Recurso e Reúso.

Os blocos existentes aparentemente eram uma classificação interna, que não era mencionada de forma explícita nos documentos consultados na subseção 5.2.2, e também não era refletida de forma tão evidente para o usuário consumidor na visualização dessas informações. Possivelmente, algumas delas não eram aparentes devido a questões de sigilo de informações privadas dos órgãos, como é o caso de dados pessoais e informações de segurança da informação.

Como apontado por Lira (2014), a qualidade e consistência da informação descritiva influencia diretamente no sucesso do consumo e da reutilização de dados abertos, consequentemente, quanto mais ricamente um determinado dado for descrito, melhor será o seu reaproveitamento em diferentes contextos. Assim, para além das **sugestões dos metadados**, apresentou-se uma sugestão de **correlação dos mesmos com os padrões DCAT-3** e, por extensão, dos padrões DC, Vcard, Foaf e Adms e as respectivas propriedades, classes e superclasses.

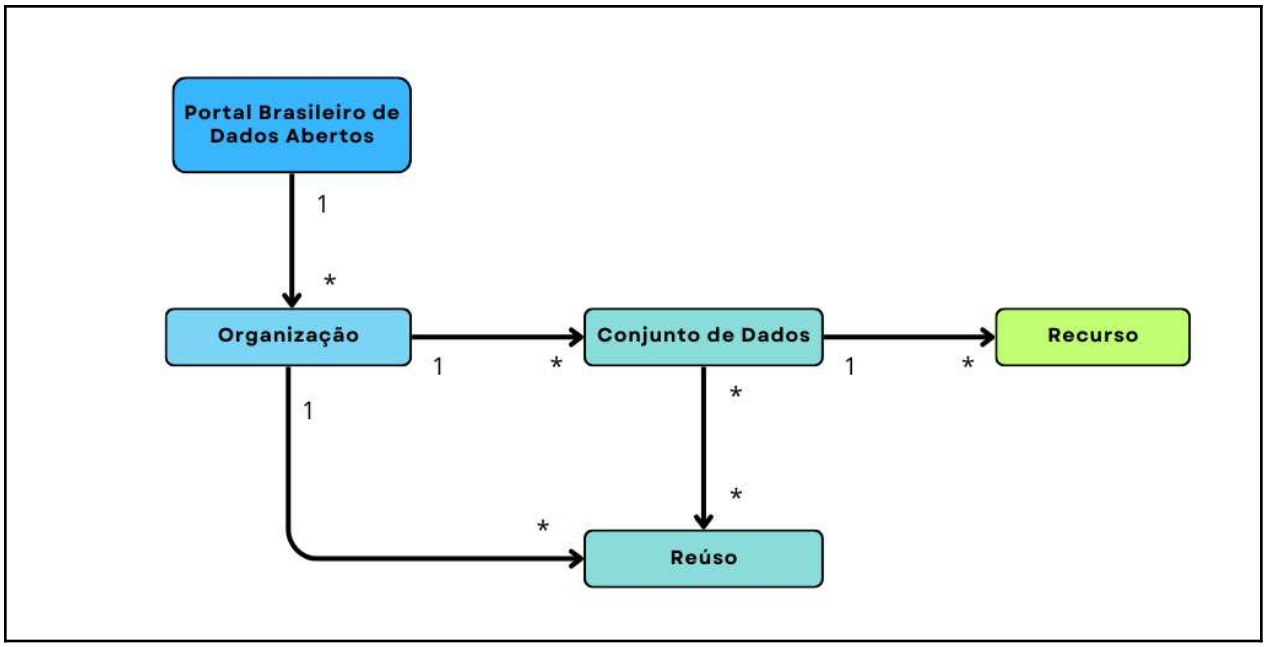
Destaca-se que a proposta apresentada nesta dissertação considerou categorias e metadados existentes no Portal, tendo em vista que o instrumento já é utilizado pelo órgãos e tem apresentado bons resultados, propondo alterações e melhorias como recomendações para a otimização desse importante instrumento.

6.1.2 Relacionamentos

A primeira movimentação da proposta foi a explicitação da estrutura de hierarquização para descrição com a divisão em categorias, considerando cada uma como uma classe que agrupa campos de descrição, sendo elas: Organização, Conjunto de Dados, Recurso e Reúso. Foram mantidas categorias identificadas no Portal que continham dados que interessavam ao cidadão, por isso, não se trabalhou com metadados da categoria Gestor de Dados, pois entende-se que esta destina-se a dados sensíveis dos órgãos, que fogem ao escopo da pesquisa e também não são visíveis para os consumidores.

Foi criada uma classe chamada Portal Brasileiro de Dados Abertos, adicionada para possibilitar a validação da proposta conceitual com a utilização do DCAT-3. Para isso, foram identificados atributos básicos para a representação do Portal, que abrange todos os demais níveis de descrição, a saber: Organização, Conjuntos de Dados, Recursos e Reúso. A partir dessa definição, foi criado um diagrama para apresentar as classes utilizadas para representar a estrutura proposta para o Portal, conforme apresentado na Figura 16.

Figura 16 - Relacionamento entre as classes propostas



Fonte: elaborada pela autora (2024).

A classe Portal Brasileiro de Dados Abertos está relacionada à Organização, uma vez que o Portal terá muitas organizações cadastradas e apresentará um relacionamento de um para muitos. Nesse sentido, a classe Organização está relacionada às classes Conjunto de dados e Reúso, sendo os relacionamentos de um para muitos, nos quais uma organização pode publicar vários conjuntos de dados ou vários reúsos. Já a classe Recurso se relaciona com a classe Conjunto de dados em um relacionamento de um para muitos, na medida em que um conjunto de dados pode ter vários recursos associados a ele.

Destaca-se que o reúso está relacionado a uma única organização, considerando a estrutura existente em relação à responsabilização e à manutenção, mas limita consumidores e outros atores ao publicarem algum reúso. Contudo, tendo em vista que ele precisa ser validado e mantido por algum órgão, essa estrutura permaneceu na recomendação. Por fim, a classe Reúso relaciona-se com a classe Conjunto de dados em um relacionamento de muitos para muitos, já que um reúso pode estar relacionado a vários conjuntos de dados e vários conjuntos de dados podem estar relacionados a muitos reúsos.

É importante destacar que todas as classes, quando representadas no banco de dados, terão um atributo identificador, id, do tipo número, criado de forma sequencial, que representará a

chave primária e permitirá identificar unicamente as entidades para fazer os relacionamentos sugeridos.

6.1.3 Metadados

A segunda movimentação da proposta foi a definição dos metadados, que considerou os requisitos da legislação, as Boas Práticas para Dados na Web, os padrões citados pelo governo e também o padrão DCAT-3, recomendado pelo W3C. Segundo Lóscio, Burle e Calegari (2019), os metadados fornecem informações complementares que ajudam os consumidores de dados a compreender melhor seu significado e sua estrutura, além de esclarecer aspectos como direitos e termos de licença, detalhes da organização responsável, qualidade dos dados, métodos de acesso e cronograma de atualização dos conjuntos de dados. Na subseção 5.2.2, que apresentou a estrutura atual do Portal, foram identificados 34 metadados visíveis para o usuário consumidor, e, na proposta atual, sugere-se uma estrutura de metadados com 84 campos visíveis para o usuário consumidor, representando um aumento de 50 metadados em relação à estrutura atual, sendo:

- Quatro referentes ao Portal Brasileiro de Dados Abertos;
- 23 referentes à Organização (manutenção de 14, adição de oito e exclusão de um);
- 27 referentes à Conjunto de Dados (manutenção de 17, adição de 10, exclusão de dois);
- 17 referentes à Recurso (manutenção de sete e adição de 10);
- 13 referentes à Reúso (manutenção de oito e adição de cinco).

Desses, sugere-se 61 metadados obrigatórios e 23 opcionais. O quadro com a estrutura completa foi apresentado no Apêndice A e para auxiliar na compreensão da proposta, utilizou-se um esquema de cores em que as linhas em fundo **Lilás** representam os metadados identificados na estrutura atual do Portal na subseção 5.2.2 e que foram mantidos, as linhas em fundo **Verde** caracterizam os metadados adicionados e as linhas em fundo **Azul** representam os metadados que sofreram alguma modificação como alteração de visibilidade ou de obrigatoriedade.

Metadados podem ser utilizados para facilitar tarefas como a descoberta e reutilização de conjuntos de dados e podem ser definidos com diferentes níveis de granularidade: desde a descrição de uma propriedade específica de um recurso até a de um conjunto de dados completo,

ou mesmo de todos os conjuntos de dados pertencentes a uma organização específica Lóscio, Burle e Calegari (2019).

Consideram-se como **obrigatórios** todos aqueles campos que são imprescindíveis para a identificação e compreensão dos dados, como título e descrição, além daqueles que necessariamente irão existir e podem inclusive serem gerados automaticamente pelo sistema de informação, como data de publicação e última alteração. Os campos **opcionais** são aqueles que contêm informações consideradas complementares, como colaborador ou atributos, aqueles que nem sempre existem necessariamente como versão anterior ou vocabulário controlado associado.

Alguns metadados têm opções restritas de preenchimento e para eles foi sugerido um controle de vocabulário para conferir conformidade e padronização, sendo eles: 'Tipo de organização', 'Natureza jurídica', 'Esfera', 'UF', 'Município', 'Licença de uso', 'Temas aos quais o dado está associado' e 'Tipo'. Essa ação possibilita um agrupamento eficaz e consistente das informações associadas a eles, pois já existem listas com valores controlados e evitam-se erros de preenchimento devido à utilização de linguagem natural, como exemplificado na subseção 5.2.3 - Tipologias de dados abertos, formatos, padrões e metadados utilizados.

Destaca-se que alguns desses campos já possuíam esse controle de vocabulário, como o caso do campo 'Temas', que utiliza o VCGE, porém, acredita-se que a explicitação da adoção dessa prática e o fornecimento de documentação com os termos controlados favoreçam a compreensão e utilização desses dados.

Davanzo e Moreira (2014) conferem aos vocabulários controlados um papel de destaque na organização de documentos em ambientes informacionais, já que eles oferecem subsídios para a recuperação precisa das informações. Nessa perspectiva, sugere-se a adoção de vocabulários controlados específicos de cada área, (por exemplo, Saúde, Educação, Transportes) para preenchimento do campo 'Assuntos/Palavras-chave', quando possível, de forma a promover maior precisão e eficácia na comunicação entre os usuários e o sistema de informações (Kobashi, 2008) e possibilitar ao usuário a identificação de conceitos que o interessem a partir de uma ferramenta normatizada e documentada. Nas subseções seguintes, serão explicitadas as categorias e os metadados propostos.

6.1.3.1 Organização

O primeiro Nível de descrição visível para o usuário consumidor refere-se à categoria **Organização**, que abrange as informações sobre o Órgão gerador/publicador dos DGA e visa atender principalmente a um requisito da LAI, que exige, em seu Art. 8º, parágrafo 1º, inciso I:

> Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.
> § 1º Na divulgação das informações a que se refere o **caput,** deverão constar, no mínimo:
> I - registro das competências e estrutura organizacional, endereços e telefones das respectivas unidades e horários de atendimento ao público (Brasil, 2011, *online*).

São propostos 23 campos de metadados, sendo 14 originários da estrutura atual do Portal, oito acréscimos e uma exclusão. Para que os campos de metadados abranjam as informações requeridas pela LAI, foram mantidos **inalterados** os campos 'Esfera', 'Órgão SIORG', 'Código SIORG', 'Poder', 'Organização Pai', 'Imagem/Logo do órgão', 'Sigla', 'CNPJ', 'UF', 'Município', 'Site da organização', 'E-mail da organização', 'Status da organização' e 'Descrição'. Esses campos foram preservados, pois contêm as informações básicas que caracterizam uma organização e sua estrutura, além de fornecerem informações de contato e de localização.

Em relação às **alterações** propostas, considerando a estrutura atual, os campos 'Tipo de organização', 'Órgão SIORG', 'Código SIORG', 'Site da organização' e 'E-mail da organização' não são visíveis para o usuário consumidor no Portal. Propõe-se que isso seja alterado, visto que são informações públicas, inclusive disponibilizadas em outras páginas do governo, como na que apresenta a estrutura do Sistema de Organização e Inovação Institucional do Governo Federal (SIORG).

Quanto ao campo 'Organização pai', ele é considerado como de preenchimento obrigatório e, analisando o contexto, considera-se a possibilidade da organização não ter uma hierarquia superior. Portanto, sugere-se que seja de preenchimento opcional, **alterando** a obrigatoriedade.

Em relação à sugestão de **exclusão**, considerou-se que os campos 'Órgão SIORG' e 'Nome do Órgão' têm o mesmo objetivo de identificar o nome do órgão para o cadastro e, por

isso, optou-se pela manutenção do primeiro, pois o segundo considera o "cadastro oficial dos órgãos e das entidades da administração direta, autárquica e fundacional do Poder Executivo Federal junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG", sendo assim, ele possui certo controle de vocabulário, que tem o objetivo de prover consistência na representação e na recuperação, evitando a dispersão de recursos informacionais (Fujita; Tolare, 2019) e facilitando a padronização e o intercâmbio de dados.

Quanto à **adição** de metadados, sugere-se a incorporação de 'Natureza jurídica', 'Logradouro', 'Número', 'Bairro', 'Complemento', 'CEP', 'Telefone de contato da organização' e 'Horário de atendimento ao público'. Esses metadados ajudam a satisfazer os requisitos da LAI, além de apresentarem uma estrutura semelhante à existente e disponível do SIORG, e esse nível de detalhamento do endereço também é encontrado no padrão GILS, mapeado no e-PMG e adotado pelo governo federal dos EUA.

O principal objetivo dos metadados da categoria Organização é fornecer informações que identifiquem e caracterizem o órgão produtor e/ou mantenedor de dados públicos. Buscou-se descrever a estrutura organizacional, como requerido pela LAI, além de informações complementares que expressam as particularidades de cada órgão. Os metadados propostos também atendem às Boas práticas para dados na web Nº 1 - Fornecer metadados e Nº 2 - Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019), possibilitando a compreensão dos dados e o conhecimento da estrutura da organização que os gerou.

Segundo Fujita (2003), a indexação que fornece a identificação de conceitos mais pertinentes ao conteúdo resulta em uma recuperação do documento mais condizente com a questão de busca e, por isso, acredita-se que a identificação detalhada e estruturada da Organização a partir desses metadados possibilita uma melhor indexação por mecanismos de busca e facilita a descoberta de dados para o usuário, na medida em que a utilização de vocabulários controlados em vários campos permite a homogeneidade no preenchimento e deixa evidentes os conceitos-chave do conteúdo e também possibilita a busca através de uma diversidade de parâmetros. A Figura 17 apresenta o compilado dos metadados sugeridos:

Figura 17 - Metadados da categoria Organização

Fonte: elaborada pela autora (2024).

O resumo da categoria Organização é apresentado no Quadro 17:

Quadro 17 - Resumo da categoria Organização

Metadado	Tipos de dados	Descrição	Nível	Tipo	Visível? (Sim/Não)	Correspondência e-PMG	Propriedade	Classe	Superclasse	Padrão	Fundamentação	Observações
Título	**String**	**Nome do catálogo**	**Portal Brasileiro de Dados Abertos**	**Obrigatório**	**Sim**	**Título (Title)**	***dcterms:title***	***dcat:Catalog***	***dcat:Dataset***	**Dublin Core**	**Boas práticas 1, 2, 3 e 5**	
Descrição	**String**	**Descrição do catálogo de dados**		**Obrigatório**	**Sim**	**description.resumo (description.abstract)**	***dcterms:description***	***dcat:Catalog***	***dcat:Dataset***	**Dublin Core**		
Homepage / Link	**URL**	**Link de acesso ao catálogo**		**Obrigatório**	**Sim**	**Localização (Location)**	***dcat:landingPage***	***dcat:Catalog***	***dcat:Dataset***	**Dcat**		
Lista de Conjuntos de Dados	**String**	**Relaciona os Conjuntos de Dados do Catálogo**		**Opcional**	**Sim**	**-**	***dcat:dataset***	***dcat:Catalog***	***dcat:Dataset***	**Dcat**		
Tipo de organização	String	Governamental: órgãos e entidades vinculados ao Governo de qualquer unidade federativa. Não Governamental: organizações de direito privado que atuam na sociedade civil e que não pertencem ao Estado.	Organização / Órgão	Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:type*	*foaf:Organization*	*foaf:Agent*	Dublin Core	LAI Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas. § 1º Na divulgação das informações a que se refere o caput, deverão constar, no mínimo: I - registro das competências e estrutura organizacional, endereços e telefones das respectivas unidades e horários de atendimento ao público;	Valores controlados: "Governamental" ou "Não governamental" Alteração de visibilidade
Natureza Jurídica	**String**	**O poder executivo federal é dividido em administração direta (ministérios) e indireta (autarquias, fundações públicas e empresas estatais)**		**Obrigatório**	**Sim**	Contexto Jurídico-administrativo (Mandate)	***dcterms:type***	***foaf:Organization***	***foaf:Agent***	**Dublin Core**		**Valores controlados: "Administração Direta" ou "Administração Indireta"**
Esfera	String	No Brasil, o sistema político está dividido em três esferas: União (Federal), Estados (Estadual) e Municípios (Municipal), além do Distrito Federal (DF)		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:type*	*foaf:Organization*	*foaf:Agent*	Dublin Core		Valores controlados: "Federal" ou "Estadual" ou "Municipal"
Órgão SIORG	String	Cadastro oficial dos órgãos e entidades da administração direta, autárquica e fundacional do Poder Executivo Federal junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:name*	*foaf:Organization*	*foaf:Agent*	Foaf		Alteração de visibilidade

Boas práticas 1, 2 e

Código SIORG	Numério	Referencial numérico que identifica o órgão ou entidade junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	dcterms:identifier	*foaf:Organization*	*foaf:Agent*	Dublin Core	29	Alteração de visibilidade
Poder	String	Executivo, legislativo, judiciário ou órgãos essenciais à justiça, como Ministérios Públicos, Defensorias Públicas e Procuradorias		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	dcterms:type	*foaf:Organization*	*foaf:Agent*	Dublin Core		Valores controlados: "Executivo" ou "Legislativo" ou "Judiciário"
Organização Pai	String	Organização à qual o órgão/entidade esteja vinculado		Opcional	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:relation*	*foaf:Organization*	*foaf:Agent*	Dublin Core		
Imagem / Logo do órgão	Arquivo de imagem	Logomarca oficial do órgão/entidade		Opcional	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:thumbnail foaf:logo*	*foaf:Organization*	*foaf:Agent*	Foaf		
Sigla	String	Abreviação que identifica o órgão		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:nick*	*foaf:Organization*	*foaf:Agent*	Foaf		
CNPJ	Numérico	Número designado pela Receita Federal para identificar o órgão/organização		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	dcterms:identifier	foaf:Organization vcard:Organization	*foaf:Agent*	Dublin Core		
UF	String	Unidade federativa em que o órgão/entidade esteja localizado		Obrigatório	Sim	-	*vcard:region*	*foaf:Organization*	*foaf:Agent*	vCard		Valores controlados
Município	String	Município em que o órgão/entidade esteja localizado		Obrigatório	Sim	-	*vcard:locality*	*foaf:Organization*	*foaf:Agent*	vCard		Valores controlados
Logradouro	**String**	**Via pública ou o local onde organização está situada**		**Obrigatório**	**Sim**	**-**	***vcard:street-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Número	**String**	**Numeração**		**Obrigatório**	**Sim**	**-**	***vcard:extend-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Bairro	**String**	**Subárea do município**		**Obrigatório**	**Sim**	**-**	***vcard:extend-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Complemento	**String**	**Informações adicionais do endereço**		**Opcional**	**Sim**	**-**	***vcard:extend-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
CEP	**Numérico**	**Código de endereçamento postal**		**Obrigatório**	**Sim**	**-**	***vcard:postal code***	***foaf:Organization***	***foaf:Agent***	**vCard**		

								n				
Site da organização	URL	Sítio eletrônico oficial do órgão/entidade		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:homepage*	*foaf:Organization*	*foaf:Agent*	Foaf		Alteração de visibilidade
Email da organização	Numérico	E-mail institucional do órgão/entidade		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:mbox vcard:hasemail*	*foaf:Organization*	*foaf:Agent*	Foaf ou vCard		Alteração de visibilidade
Telefone de contato da organizaçã o	**Numérico**	**Telefone de contato do órgão/entidade**		**Obrigatório**	**Sim**	**Contexto Jurídico-administrativo (Mandate)**	***vcard:tel***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Horário de atendimento ao público	**String**	**Escala de dias e horários de atendimento ao público**		**Obrigatório**	**Sim**	**-**	***vcard:hours***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Status da organização	String	Ativo: o órgão/entidade está ativo junto à Receita Federal do Brasil. Extinto: órgão/entidade não executa mais atividades e estão inativos junto à Receita Federal do Brasil.		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:type*	*foaf:Organization*	*foaf:Agent*	Dublin Core		Valores controlados: "Ativo" ou "Extinto"
Descrição (da organização)	String	Descrição do órgão/entidade, como área de atuação, missão e demais informações que o órgão/entidade achar pertinente compartilhar		Obrigatório	Sim	description.resumo (description.abstract)	*dcterms:description*	*foaf:Organization*	*foaf:Agent*	Dublin Core		

Fonte: elaborado pela autora (2024).

6.1.3.2 Conjunto de Dados

A categoria **Conjunto de Dados** abrange uma coleção de dados que são publicados ou mantidos por uma organização e estão disponíveis para acesso ou *download* em um ou mais formatos (Albertoni *et al,* 2024). Alguns exemplos de conjuntos de dados são uma coleção de tabelas relacionadas entre si, dados extraídos de um mesmo sistema de informações ou uma API de dados abertos (Brasil, s.d.f).

Os Conjuntos de Dados são considerados os elementos principais retornados a partir das buscas no Portal e cada um deles possui pelo menos um recurso associado. (Brasil, s.d.d). Nesta categoria, são 27 metadados, sendo 17 da estrutura atual do Portal, 10 acréscimos e duas exclusões.

Conforme a Boa prática para dados na web Nº 2 - Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019), listada no Quadro 9, são sugeridos os seguintes metadados para a descrição dos Conjuntos de Dados: ‘Título’, ‘Descrição’, ‘Palavras-chave’, ‘Data de publicação’, ‘Entidade responsável (publicadora de dados)’, ‘Ponto de contato’, ‘Cobertura Espacial’, ‘Período temporal’, ‘Data da última modificação’ e ‘Temas/categorias’. Destes, apenas o campo ‘Data da última modificação’ não existia na estrutura atual e foi sugerida sua **adição** com a nomenclatura de ‘Última alteração’.

Além dos campos compatíveis com a Boa prática para dados na web Nº 2 - Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019), foram mantidos **inalterados** os campos ‘Item de inventário de origem’, que atende à Boa prática para dados na web Nº 3 - Fornecer metadados estruturais (Lóscio; Burle; Calegari, 2019), ‘Licença de uso’, relacionado à Boa prática para dados na web Nº 4 - Fornecer informações sobre a licença de dados (Lóscio; Burle; Calegari, 2019), e ‘Periodicidade de atualização’, correspondente à Boa prática para dados na web Nº 21 - Fornecer dados atualizados (Lóscio; Burle; Calegari, 2019).

No caso dos metadados ‘Área técnica responsável pelo dado’, ‘E-mail da área técnica responsável’, ‘Cobertura temporal início’, ‘Cobertura temporal fim’, ‘Descontinuado’, ‘Data de descontinuação’ e ‘Conjuntos de dados associados (interoperabilidade)’, a presença deles foi observada apenas no manual para publicadores de dados, mas, aparentemente, não são visíveis para o usuário consumidor, por isso, sugere-se uma **alteração** de visibilidade. Como não são

campos que contêm informações sigilosas ou dados sensíveis, considera-se que deve ser aplicado o princípio de transparência da publicidade como regra, e o sigilo, como exceção.

Em relação aos metadados ‘Área técnica responsável pelo dado’ e ‘E-mail da área técnica responsável pelo dado’, sugere-se a **alteração** de visibilidade para abarcar uma importante vertente apontada por Lira (2014), mas não contemplada nessa proposta: a disponibilização de informações que descrevem os metadados. Conforme o autor, devido às dificuldades para a criação e disponibilização de metadados para a descrição de dados abertos, deseja-se também a promoção do reúso de metadados para evitar redundância e colaborar com a padronização da informação e, portanto, é fundamental fornecer, além de metadados que descrevam os dados, informações que os descrevam.

Levando essa perspectiva em consideração, espera-se que esses metadados sirvam como um ponto de contato para dúvidas sobre os dados e os metadados que os descrevem caso as informações disponibilizadas sejam insuficientes para satisfazer as necessidades de informação do usuário. Relaciona-se esses campos também às Boas práticas para dados na web Nº 1 - Fornecer metadados, Nº 2 - Fornecer metadados descritivos, Nº 5 - Fornecer informações de procedência de dados e Nº 29 - Coletar feedback de consumidores de dados do (Lóscio; Burle; Calegari, 2019).

Para os campos ‘Cobertura espacial’ e ‘Granularidade espacial’, sugere-se a **alteração** de visibilidade, pelo mesmo motivo relatado acima, e também a **alteração** da obrigatoriedade, originalmente campos opcionais, que sugere-se que sejam obrigatórios, pois entende-se que, sendo dados gerados por Organizações vinculadas ao Governo Federal, a cobertura e granularidade espacial sempre existirão. Esses metadados aderem à Boa prática para dados na web Nº 13 - Usar representações de dados que sejam independentes de localidade (Lóscio; Burle; Calegari, 2019) e, quando isso não for possível, fornecer metadados sobre os parâmetros regionais dos conteúdos de dados.

Já os metadados ‘Descontinuado’ e ‘Data de descontinuação’ podem associar-se à Boa prática para dados na web Nº 22 - Fornecer uma explicação para os dados que não estão disponíveis (Lóscio; Burle; Calegari, 2019), para que os consumidores saibam que os dados referidos, extraídos do conjunto de dados corrente, não estão disponíveis ou podem ser disponibilizados sob condições diferentes. Por fim, sugere-se a alteração de visibilidade para o

metadado ‘Conjuntos de dados associados (interoperabilidade)’, que atende às Boas Práticas para Dados na Web Nº 2, 3 e 5 (Lóscio; Burle; Calegari, 2019), já citadas acima.

Em relação às **exclusões**, sugeriu-se a exclusão desta categoria dos metadados ‘Versão’ e ‘Atualização de versão’, pois entende-se que é mais vantajoso relacioná-los a um recurso, pois um conjunto de dados pode ter vários recursos e cada recurso pode ter diferentes versões.

Sugere-se a **adição** dos campos opcionais ‘Colaborador’, ‘API habilitada’, ‘Vocabulário controlado associado', ‘Lista dos recursos cadastrados’, ‘Documentação’, ‘Granularidade temporal’ e ‘Metodologia’; do campo obrigatório ‘Identificador’ e dos campos de geração automática ‘Data de publicação’ e ‘Última alteração’.

Em relação ao campo ‘Colaborador’, acredita-se que ele seja adequado em um contexto para a promoção da interoperabilidade e do reúso de dados, onde pode haver intercâmbio de informações e, por isso, é interessante registrar possível cooperação para manter um histórico. A Boa prática para dados na web Nº 5 - Fornecer informações de procedência dos dados (Lóscio; Burle; Calegari, 2019), alerta justamente sobre isso e justifica que a procedência é uma das formas que os consumidores têm para julgar a qualidade de um conjunto de dados e que a origem e história ajudam a determinar a confiabilidade do dado e a fornecer contexto para interpretação. Além disso, é um campo presente no e-PMG e mapeado também nos padrões AGLS, e-GMS e MIP.

Já o campo ‘API habilitada’ atende à Boa prática para dados na web Nº 23 - Disponibilizar dados por meio de uma API (Lóscio; Burle; Calegari, 2019), que garante aos consumidores maior flexibilidade e facilidade de processamento dos dados, e também à Boa prática para dados na web Nº 35 - Citar a publicação original do conjunto de dados (Lóscio; Burle; Calegari, 2019). Ainda que a utilização de uma API exija certo conhecimento técnico por parte do usuário, a disponibilização de dados por meio dela pode ser o incentivo para a criação de aplicações que atendam a usuários com todos os níveis de habilidade informacional.

O campo ‘Vocabulário controlado associado’ atende à Boa prática para dados na web Nº 15 - Reutilizar vocabulários (Lóscio; Burle; Calegari, 2019), dando preferência aos padronizados, e pode também complementar os dicionários de dados que porventura estejam associados a algum vocabulário controlado. Segundo Davanzo e Moreira:

> O vocabulário controlado agrega valor ao processo de organização e recuperação da informação porque sua utilização possibilitará o controle do repertório terminológico que será utilizado para descrever e/ou representar os conceitos que

> descrevem um determinado campo. Deve-se ressaltar, portanto, que a adoção de uma linguagem clara, padronizada e, na medida do possível, comum a todos diminuirá consideravelmente as dificuldades de comunicação que efetivamente existem entre usuários e sistema (2014, p. 4).

Em relação aos metadados ‘Documentação’, ‘Granularidade temporal’ e ‘Metodologia’, eles constam na Cartilha Técnica para Publicação de Dados Abertos, que é uma publicação do próprio governo e está indicada como documento norteador para a publicação de dados no Portal. Sinaliza-se que o padrão de metadados constante nesta cartilha é o instrumento a ser seguido e, por isso, sugere-se a adição desses campos como opcionais. Entende-se que eles complementam a proposta, atendendo à Boa prática para dados na web Nº 2 - Fornecer metadados descritivos do (Lóscio; Burle; Calegari, 2019).

O campo ‘Identificador’ atende à Boa prática para dados na web Nº 9 - Usar URIs persistentes como identificadores de conjuntos de dados (Lóscio; Burle; Calegari, 2019), que permite a identificação básica dos dados e é pré-condição essencial para o gerenciamento e a reutilização dos dados de forma adequada. Dessa forma, os conjuntos de dados ou as informações sobre os conjuntos de dados serão encontrados e citados com facilidade em qualquer momento, independentemente do *status*, da disponibilidade ou do formato desses dados (Lóscio; Burle; Calegari, 2019).

Por fim, os metadados ‘Data de publicação’ e ‘Última alteração’ atendem à Boa prática para dados na web Nº 21 - Fornecer dados atualizados (Lóscio; Burle; Calegari, 2019), e auxiliam o consumidor de dados a conhecer um pouco do histórico do conjunto de dados acessado. A Figura 18 apresenta o agrupamento dos metadados da proposta:

Figura 18 - Metadados da classe Conjunto de Dados

Fonte: elaborado pela autora (2024).

O resumo da categoria Conjuntos de Dados é apresentado no Quadro 18:

Quadro 18 - Resumo da categoria Conjuntos de Dados

Metadado	Tipos de dados	Descrição	Nível	Tipo	Visível? (Sim/Não)	Correspondência e-PMG	Propriedade	Classe	Superclasse	Padrão	Fundamentação	Observações
Título do conjunto de dados	String	Nome do conjunto de dados. A nomenclatura deve ser idêntica à que foi elencada no Plano de Dados Abertos – PDA, quando esse instrumento for aplicável e o conjunto tiver sido previsto no cronograma daquele documento	Conjunto de dados	Obrigatório	Sim	Título (Title)	dcterms:title	dcat:Dataset1	dctype:Dataset	Dublin Core	Boas práticas 1 e 2	
Organização publicadora	String	Campo que identifica a organização que está cadastrando o conjunto de dados		Obrigatório	Sim	criador.autor (creator.autor)	dcterms:publisher	dcat:Dataset1	dctype:Dataset	DublinCore	Boas práticas 1, 2 e 5	
Item de inventário de origem	String	Indicar a qual item do inventário o conjunto de dados está relacionado		Opcional	Sim	Fonte (Source)	dcterms:source	dcat:Dataset1	dctype:Dataset	DublinCore	Boas práticas 1, 3 e 5	
Descrição do conjunto de dados	String	Descrição sobre o conjunto de dados, incluindo metadados que o identifique		Obrigatório	Sim	description.resumo (description.abstract)	dcterms:description	dcat:Dataset1	dctype:Dataset	DublinCore	Boas práticas 1 e 2	
Licença de uso	String	Selecionar sob qual espécie de licença o conjunto de dados será publicado		Obrigatório	Sim	Direitos (Rights)	dcterms:license	dcat:Dataset1	dctype:Dataset	DublinCore	Boas práticas 1 e 4	Valores controlados
Área técnica responsável pelo dado	String	Indicar qual área técnica do órgão/entidade é responsável pelo conjunto de dados		Obrigatório	Sim	criador.produtor (creator.productor)	dcterms:creator	dcat:Dataset1	dctype:Dataset	DublinCore	Boas prátocas 1, 2 e 5	Alteração de visibilidade
Email da área técnica responsável	String	Indicar e-mail de contato da área técnica responsável pelo conjunto de dados		Obrigatório	Sim	-	dcat:contactPoint	dcat:Dataset1	dctype:Dataset	Dcat	Boas práticas 1, 2 e 29	Alteração de visibilidade
Periodicidade de atualização	String	Indicar qual será a periodicidade com que o conjunto de dados será atualizado		Obrigatório	Sim	data.frequenciaDeAtualização	dcterms:accrualPeriodicity	dcat:Dataset1	dctype:Dataset	Dublin Core	Boas práticas 1 e 21	
Temas aos quais o dado está associado	String	Escolha a área temática à qual o conjunto de dados esteja vinculado		Obrigatório	Sim	Assunto.categoria (subject.category)	dcat:themeTaxonomy	dcat:Dataset1	dctype:Dataset	Dcat	Boas práticas 1 e 2	Valores controlados (VCGE)
Assuntos / Palavras-chave	String	Resumo dos principais aspectos do conjunto de dados		Obrigatório	Sim	assunto.palavra-chave	dcat:keyword	dcat:Dataset1	dctype:Dataset	Dcat	Boas práticas 1 e 2	

Cobertura temporal início	Numérico	Data ou período de início dos dados disponibilizados		Obrigatório	Sim	abrangência.temporal	dcat:startDate	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1 e 2	Alteração de visibilidade
Cobertura temporal fim	Numérico	Data ou período de fim dos dados disponibilizados		Opcional	Sim	abrangência.temporal	dcat:endDate	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1 e 21	Alteração de visibilidade
Cobertura espacial	String	Esfera de cobertura espacial do conjunto de dados		Obrigatório	Sim	abrangência.espacial	dcterms:spatial	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 13	Alteração da obrigatoriedade e visibilidade
Granularidade espacial	String	Nível de granularidade espacial do conjunto de dados		Obrigatório	Sim	abrangência.espacial	dcterms:spatial	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 2	Alteração da obrigatoriedade e visibilidade
Descontinuado	String	Indicar se conjunto de dados foi descontinuado e não passará por futuras atualizações		Opcional	Sim	-	*adms:versionNotes*	*dcat:Dataset1*	*dctype:Dataset*	Adms	Boas práticas 1 e 22	Alteração de visibilidade
Data descontinuação	String	Data a partir da qual o conjunto de dados foi descontinuado		Opcional	Sim	data.disponibilidade (date.available) data.validade (date.valid)	*dcterms:valid*	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 22	Alteração de visibilidade
Conjuntos de dados associados (interoperabilidade)	String	Indicar outros conjuntos de dados que estejam associados ao conjunto de dados cadastrado		Opcional	Sim	Relação (Relation)	dcat:qualifiedRelation dcat:dataset *dcterms:relation*	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core ou Dcat	Boas práticas 1 , 3 e 31	Alteração de visibilidade
Colaborador	**String**	**Pessoa ou entidade que contribuiu para a criação, revisão ou manutenção do recurso**		**Opcional**	**Sim**	**Colaborador (Contributor)**	***dcterms:contributor***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 5**	
API habilitada	**String**	**Indica se o recurso possui uma interface de programação (API) disponível para interação automatizada**		**Opcional**	**Sim**	**-**	***dcat:endpointURL***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1, 23 e 35**	
Identificador	**URI / IRI**	**Código ou número único que identifica o recurso ou registro de forma exclusiva**		**Obrigatório**	**Sim**	**Identificador (Identifier)**	***dcterms:identifier***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1 e 9**	
Vocabulário contorlado associado	**String**	**Conjunto padronizado de termos utilizados para descrever ou categorizar os dados**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 15**	
Lista dos Recursos cadastrados	**String**	**Lista os recursos que estão cadastrados no conjunto de dados**		**Opcional**	**Sim**	**-**	***dcat:distribution***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 3**	

Data de publicação	**String**	**Data em que o Conjunto de Dados foi publicado no Portal**		**Automático**	**Sim**	**data.publicação**	***dcterms:issued***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Última alteração	**String**	**Última alteração do conjunto de dados**		**Automático**	**Sim**	**data.alteração**	***dcterms:modified***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Documentação	**String**	**URL de documento que expõe detalhes sobre o conjunto de dados**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 2**	
Granularidade temporal	**String**	**Precisão temporal da cobertura temporal.**		**Opcional**	**Sim**	**abrangência.temporal**	***dcterms:temporal***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1 e 2**	
Metodologia de criação	**String**	**Processo de criação dos dados**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 2**	

Fonte: elaborado pela autora (2024).

6.1.3.3 Recursos

Quanto à categoria **Recurso**, compreende-se que cada recurso representa uma fonte de dados, podendo ser um arquivo, como uma planilha, um método de *webservice* / API ou um documento. Segundo informações do Portal, cada recurso poderia ser catalogado como um conjunto de dados separado, entretanto, é aconselhável que recursos sejam agrupados quando forem oriundos da mesma base de dados ou possuírem metadados em comum, de forma a facilitar a busca e a compreensão do conteúdo (Brasil, s.d.f). Assim, os recursos estão sempre associados a um conjunto de dados, ainda que eles mesmos possam ser conjuntos de dados. Eles correspondem a 17 metadados totais na proposta, sendo sete originários da estrutura atual do Portal e 10 adições.

Conforme a Boa prática para dados na web Nº 2 - Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019), os metadados descritivos podem incluir as seguintes características genéricas de uma distribuição: ‘Título’, ‘Descrição’, ‘Data da publicação’ e ‘Formato de mídia’. Uma distribuição representa um determinado formato disponível de um conjunto de dados, como, por exemplo, um arquivo no formato CSV para *download* ou um API (Albertoni *et al*, 2024). O único campo dentre esses que não estava contemplado originalmente era ‘Data da publicação’, e sugere-se a sua **adição**. Além desses, foram mantidos **inalterados** os campos ‘URL do recurso’, ‘Tipo’, ‘Data de criação’ e ‘Última alteração’.

Relaciona-se os campos ‘Formato de mídia’ e ‘Tipo’ também a Boa prática para dados na web Nº 3 - Fornecer metadados estruturais (Lóscio; Burle; Calegari, 2019), os campos ‘Data da publicação’, ‘Data de criação’ e ‘Última alteração’ a Boa prática para dados na web Nº 21 - Fornecer metadados atualizados (Lóscio; Burle; Calegari, 2019) e o metadados ‘URL do recurso’ à Boa prática para dados na web Nº 5 - Fornecer informações de procedência de dados e Nº 20 - Fornecer acesso em tempo real (Lóscio; Burle; Calegari, 2019).

Sugere-se a **adição** dos campos ‘Versão’ e ‘Atualização de versão’, que eram originários da categoria Conjuntos de Dados, pois entende-se que cada recurso de um conjunto de dados pode ter sua própria versão. Como complemento, sugere-se também a adição dos campos opcionais ‘Notas de versão’ e ‘Versão anterior’. Essas adições atendem à Boa prática para dados na web Nº 7 - Fornecer indicador de versão e à Boa prática para dados na web Nº 8 - Fornecer o histórico de versão e o bom versionamento (Lóscio; Burle; Calegari, 2019), já que proporcionam

aos consumidores compreender se uma nova versão está disponível e permitem comparações. Além desses, sugere-se a **adição** dos campos ‘Assuntos/Palavras-chave’, ‘Idioma’, ‘Tamanho do arquivo’, ‘Sigilo/Direitos de acesso’, ‘Normativa/Observância legal’.

O campo ‘Assuntos/Palavras-chave’ é um descritor básico que visa realizar uma indexação do conteúdo do recurso de forma resumida, com a utilização de termos relevantes[2] . Esse é um metadado que atende à Boa prática para dados na web Nº 2 - Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019) e, como sugerido para a categoria Conjuntos de Dados, a utilização de um vocabulário controlado para preenchimento poderia aprimorar a representação descritiva ao padronizar também o conteúdo.

Ainda que o Portal abranja dados produzidos por órgãos brasileiros, entende-se que não existe a obrigatoriedade de esses dados estarem no idioma Português e, por isso, vê-se a necessidade do campo ‘Idioma’ para registrar essa especificação.

O campo ‘Sigilo/Direitos de acesso’ estabelece restrições ou permissões de acesso, delimitando quem pode visualizar ou acessar um recurso, e é complementado pelo campo ‘Normativa/Observância legal’, que busca esclarecer algum tipo de restrição que possa haver na utilização de um recurso devido, por exemplo, a alguma legislação, como a LGPD, ou como a previsibilidade na LAI de restrição temporária de acesso público em questões que envolvam a segurança da sociedade e do Estado. A Figura 19 apresenta os metadados propostos para a categoria Recursos:

Figura 19 - Metadados da categoria Recursos

Fonte: elaborada pela autora (2024).

O resumo da categoria Recursos é apresentado no Quadro 19:

Quadro 19 - Resumo da categoria Recursos

Metadado	Tipos de dados	Descrição	Nível	Tipo	Visível? (Sim/Não)	Correspondência e-PMG	Propriedade	Classe	Superclasse	Padrão	Fundamentação	Observações
Título de recurso	String	Nome do recurso	Recurso (dados, documentação, dicionário de dados, API ou outros)	Obrigatório	Sim	Título (Title)	dcterms:title	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1 e 2	
URL do recurso	URL	Endereço da web onde o recurso está hospedado		Obrigatório	Sim	Localização (Location)	*dcat:accessURL*	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1, 5 e 20	
Tipo do recurso	String	Selecionar se o recurso se refere a dados, documentação, dicionário de dados, API ou outros		Obrigatório	Sim	Tipo (Type)	dcterms:type	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1 e 3	Valores controlados
Formato do recurso	String	Selecionar sob qual formato o recurso será disponibilizado		Obrigatório	Sim	Formato (Format)	*dcterms:format dcat:mediaType*	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core ou Dcat	Boas práticas 1 e 3	
Descrição do recurso	String	Breve descrição sobre o recurso		Obrigatório	Sim	description.resumo (description.abstract)	dcterms:description	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1 e 2	
Data de criação do recurso	String	Data de criação do recurso		Obrigatório	Sim	**data.criação**	dcat:startDate	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1, 2 e 21	
Data de publicação	**String**	**Data em que o recurso foi publicado no Portal**		**Automático**	**Sim**	**data.publicação**	***dcterms:issued***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Última alteração	String	Última alteração do recurso		Automático	Sim	data.alteração	dcterms:modified	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1, 2 e 21	
Versão	Numérico	Versão do conjunto de dados que está sendo cadastrada		Opcional	Sim	Identificador (Identifier) identificador.versão	dcat:hasCurrentVersion	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1 e 7	Mudança de Conjuntos de Dados para Recurso
Atualização de versão	Numérico	Indicar se houve atualização da versão do conjunto de dados disponibilizado na última alteração realizada		Opcional	Sim	-	dcat:hasVersion	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1 e 8	Mudança de Conjuntos de Dados para Recurso
Notas de versão	**String**	**Observações sobre atualizações ou alterações na versão**		**Opcional**	**Sim**	**-**	***adms:versionNotes***	***dcat:Distribution***	***dcat:Dataset1***	**Adms**	**Boas práticas 1 e 8**	
Versão anterior	**Numério**	**Indica se o recurso tem versão prévia**		**Opcional**	**Sim**	**-**	***dcat:previousVersion***	***dcat:Distribution***	***dcat:Dataset1***	**Dcat**	**Boas práticas 1 e 8**	
Assuntos / Palavras-ch	**String**	**Resumo dos principais aspectos**		**Obrigatório**	**Sim**	**Assunto (Subject)**	***dcat:keyword***	***dcat:Distributi***	***dcat:Dataset1***	**Dcat**	**Boas práticas 1 e 2**	

ave		**do recurso**						***on***				
Idioma	**String**	**Língua na qual o conteúdo do recurso está escrito ou apresentado**		**Obrigatório**	**Sim**	**Idioma (Language)**	***dcterms:language***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1 e 2**	
Tamanho do arquivo	**Numérico**	**Medida do tamanho do arquivo ou conjunto de dados em bytes**		**Obrigatório**	**Sim**	**format.dimensão**	***dcat:byteSize***	***dcat:Distribution***	***dcat:Dataset1***	**Dcat**	**Boas práticas 1 e 3**	
Sigilo / Direitos de acesso	**String**	**Nível de restrição ou permissões de acesso, indicando quem pode visualizá-lo ou utilizá-lo**		**Opcional**	**Sim**	**Direitos (Rights)**	***dcterms:accessRights***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1 e 2**	
Normativa / Observância legal	**String**	**Normativas / Legislação aplicável**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1 e 2**	

Fonte: elaborado pela autora (2024).

6.1.3.4 Reúso

Por fim, tem-se a categoria **Reúso**, que refere-se às aplicações criadas a partir de conjuntos de dados ou recursos já existentes, podendo ser aplicativos, sites, entre outros. Este nível possui 13 metadados, sendo oito da atual estrutura e cinco adições. Mantiveram-se **inalterados** os campos 'Nome', 'URL', 'Descrição', 'Tipo', 'Imagem' e 'Organização a qual o recurso está associado'. Os metadados 'Nome', 'Descrição' e 'Imagem' associam-se à Boa prática para dados na web Nº 2 - Fornecer metadados descritivos (Lóscio; Burle; Calegari, 2019), o campo 'URL' relaciona-se com as Boas práticas para dados na web Nº 5 - Fornecer informações sobre procedência dos dados e Nº 20 - Fornecer acesso em tempo real (Lóscio; Burle; Calegari, 2019), o metadado 'Tipo' se vincula à Boa prática para dados na web Nº 3 - Fornecer metadados estruturais (Lóscio; Burle; Calegari, 2019), e o metadado 'Organização a qual o recurso está associado' atende às citadas Boas práticas para dados na web Nº 3 e Nº 5 (Lóscio; Burle; Calegari, 2019).

Sugere-se a **alteração** de visibilidade para o campo 'Conjuntos de dados de origem', para que o usuário consumidor possa verificar a correlação entre os conjuntos de dados, caso assim desejar. Esse metadado atende às Boas práticas para dados na web Nº 3 - Fornecer metadados estruturais, Nº 5 - Fornecer informações de procedência dos dados, Nº 31 - Enriquecer dados por meio da geração de novos dados e Nº 35 - Citar a publicação original do conjunto de dados (Lóscio; Burle; Calegari, 2019).

Propõe-se também a **alteração** do campo 'Nome do responsável' para 'Área técnica responsável', de forma a estabelecer um ponto de contato mais administrativo e menos pessoal, preservando informações sensíveis/sigilosas. Para complementar esse ponto de contato, sugere-se a **adição** do campo 'E-mail da área técnica responsável'. Esses metadados atendem à Boa prática para dados na web Nº 2 - Fornecer metadados descritivos e podem, também, atender à Boa prática para dados na web Nº 29 - Coletar feedback de consumidores de dados (Lóscio; Burle; Calegari, 2019), já que fornecem um ponto de contato.

Sugere-se, ainda, a adição dos metadados 'Data de criação do reúso', 'Data de publicação do reúso' e 'Última alteração', sendo o primeiro de preenchimento obrigatório e os demais, de geração automática pelo sistema, com intenção de fornecer mais detalhamento sobre o histórico

do reúso e atender à Boa prática para dados na web Nº 21 - Fornecer dados atualizados (Lóscio; Burle; Calegari, 2019).

Essa categoria é a que possui a menor quantidade de elementos cadastrados no Portal, contando com apenas 10 reúsos cadastros até 18/11/24, segundo informações da página inicial. Uma hipótese para esse fato é a de que os usuários ainda não se apropriam dos dados de forma a aproveitarem todo o seu potencial e, por isso, ocorre a baixa expressividade de reúso dos dados no Portal ou daqueles que utilizam, porém não cadastram esse reúso como novo recurso no Portal. Soma-se a isso a circunstância de que o reúso se vale de conjuntos de dados ou recursos já cadastrados, quando parte da descrição já foi realizada, sendo necessários menos metadados para a identificação e especificação dessa categoria.

Dos 13 metadados listados, apenas ‘Imagem do reúso’ e ‘E-mail da área técnica responsável’ não têm correspondência no e-PMG, sendo que o campo ‘URL do reúso’, correlacionado com a propriedade ‘Localização (*Location*)’, é considerado um dos mais importantes, porque informações mais detalhadas e específicas podem estar disponíveis no local onde o reúso está hospedado. A Figura 20 apresenta os metadados propostos para a categoria Reúso:

Figura 20 - Metadados da categoria Reúso

Fonte: elaborada pela autora (2024).

O resumo da categoria Reúso é apresentado no Quadro 20:

Quadro 20 - Resumo da categoria Reúso

Metadado	Tipos de dados	Descrição	Nível	Tipo	Visível? (Sim/Não)	Correspondência-PMG	Propriedade	Classe	Superclasse	Padrão	Fundamentação	Observações
Título do reúso	String	Nome do reúso por extenso	Reúso	Obrigatório	Sim	Título (Title)	dcterms:title	dcat:Dataset2	dctype:Dataset	Dublin Core	Boas práticas 1 e 2	
URL do reúso	URL	Endereço da web onde o reúso está hospedado		Obrigatório	Sim	Localização (Location)	dcat:accessURL	dcat:Dataset2	dctype:Dataset	Dcat	Boas práticas 1 e 5	
Descrição do reúso	String	Explicação sobre o reúso		Obrigatório	Sim	description.resumo (description.abstract)	dcterms:description	dcat:Dataset2	dctype:Dataset	Dublin Core	Boas práticas 1 e 2	
Tipo do reúso	String	Selecionar entre as opções fornecidas – aplicativo, site, dados e outros)		Obrigatório	Sim	Tipo (Type)	dcterms:type	dcat:Dataset2	dctype:Dataset	Dublin Core	Boas práticas 1 e 3	Valores controlados
Imagem do reúso	Arquivo de imagem	Imagem referente ao reúso		Opcional	Sim	-	foaf:thumbnail	dcat:Dataset2	dctype:Dataset	Foaf	Boas práticas 1 e 2	
Organização a qual o reúso está associado	String	Selecionar qual órgão/entidade a que o reúso está relacionado		Obrigatório	Sim	Relação (Relation) ou Fonte (Source)	dcterms:relation	dcat:Dataset2	dctype:Dataset	Dublin Core	Boas práticas 1 e 5	
Conjuntos de dados de origem	String	Indicar o(s) conjunto(s) de dados que foram utilizados para o reúso		Obrigatório	Sim	Relação (Relation) ou Fonte (Source)	dcat:Dataset	dcat:Dataset2	dctype:Dataset	Dcat	Boas práticas 1, 5, 31 e 35	
Assuntos / Palavras-chave	String	Palavras que destacam o tema do reúso		Obrigatório	Sim	Assunto (Subject)	dcat:keyword	dcat:Dataset2	dctype:Dataset	Dcat	Boas práticas 1 e 2	
Área técnica responsável pelo reúso	**String**	**Setor ou equipe encarregada pela produção ou manutenção do reúso**		**Obrigatório**	**Sim**	**criador.produtor (creator.producto r)**	**dcterms:creator**	**dcat:Dataset2**	**dctype:Dataset**	**Dublin Core**	**Boas práticas 1 e 5**	**Substituição de "Nome do responsável" por "Área técnica responsável"**
Email da área técnica responsável	**String**	**Endereço de e-mail de contato da área responsável pelo reúso**		**Obrigatório**	**Sim**	**-**	**dcat:contactPoint**	**dcat:Dataset2**	**dctype:Dataset**	**Dcat**	**Boas práticas 1 e 2**	
Data de criação do reúso	**String**	**Data de criação do reúso**		**Obrigatório**	**Sim**	**data.criação**	**dcat:startDate**	**dcat:Dataset2**	**dctype:Dataset**	**Dcat**	**Boas práticas 1, 2 e 21**	
Data de publicação	**String**	**Data em que o recurso foi publicado no Portal**		**Automático**	**Sim**	**data.publicação**	**dcterms:issued**	**dcat:Dataset2**	**dctype:Dataset**	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Última alteração	**String**	**Última alteração do reúso**		**Automático**	**Sim**	**data.alteração**	**dcterms:modified**	**dcat:Dataset2**	**dctype:Dataset**	**Dublin Core**	**Boas práticas 1, 2 e 21**	

Fonte: elaborado pela autora (2024).

Considerando a proposta geral, sempre que possível, foi sugerida a preservação da estrutura atual, de forma a gerar menor impacto e maior simplicidade em uma possível implementação. As sugestões de alterações e adições foram pautadas buscando atender às legislações pertinentes em sua totalidade e pensando no usuário consumidor, que se beneficiará quanto mais ricas e detalhadas forem as informações fornecidas. A proposta completa pode ser visualizada no Apêndice A.

6.1.4 Correspondência com o padrão DCAT-3

Segundo Lóscio, Burle e Calegari (2019), uma boa prática genérica para a publicação de dados na Web é o uso de padrões e, para Tomoyose (2021), os vocabulários podem ser utilizados para padronizar os termos de metadados empregados. Após a definição dos metadados propostos para representação no Portal, cada um foi correlacionado com uma propriedade e uma classe utilizando o vocabulário DCAT-3, que possui diversas características que o adequam ao contexto, como a possibilidade de utilização de outros vocabulários, aumentando a interoperabilidade dos dados e a possibilidade de disponibilização de uma versão legível por máquina dos metadados descritivos.

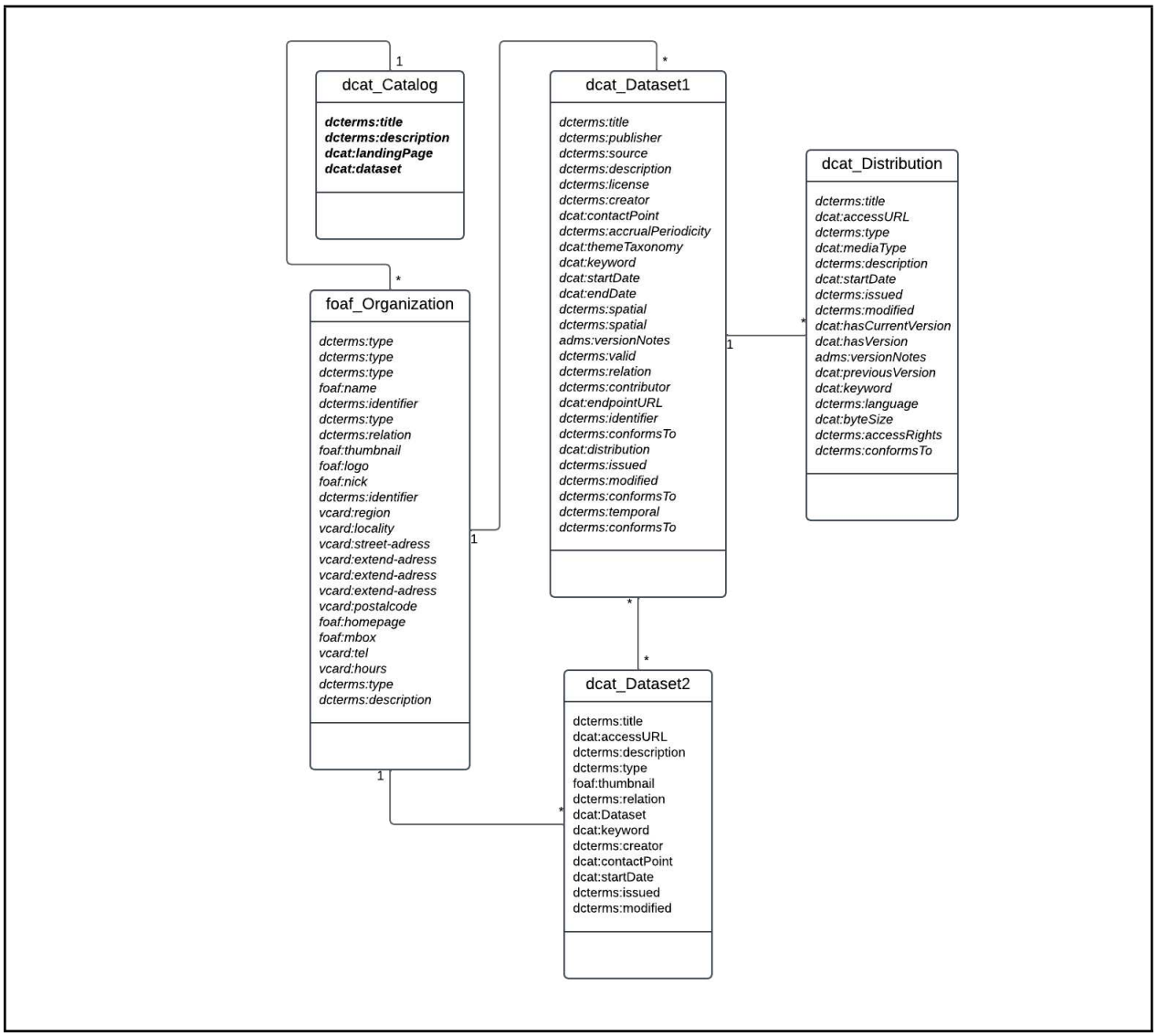
Para a validação da estrutura proposta, após a correlação dos metadados com propriedades do DCAT, foi criada uma sintaxe em Turtle, com a descrição de classes e propriedades, cujo código foi validado em uma ferramenta gratuita chamada *Turtle Validator*[1]. Na prática, isso significa que o código está corretamente descrito e, no contexto do trabalho, representa o primeiro passo para corroborar o esquema proposto.

Das sete classes principais do DCAT-3, foram utilizadas três para modelar a proposta de metadados de acordo com esse padrão, sendo elas: **dcat:Catalog**, representando o Portal Brasileiro de Dados Abertos, o catálogo central de disponibilização de dados da administração pública federal; **dcat:Dataset1**, representando os Conjuntos de dados; **dcat:Dataset2**, representando os Reúsos publicados no Portal, e **dcat:Distribution** para abranger os metadados que caracterizam cada versão de um Recurso associado a um conjunto de dados. Como complemento, utilizou-se a classe **foaf:Organization**, indicada pela própria documentação do DCAT-3 para a descrição das Organizações publicadoras.

[1] http://ttl.summerofcode.be/

A partir dessa definição, foi criado um diagrama para apresentar as classes e propriedades utilizadas para representar a estrutura proposta para o Portal. O relacionamento entre as classes significa que o Portal Brasileiro de Dados Abertos (dcat:Catalog) pode ter inúmeras organizações publicadoras (foaf:Organization) cadastradas e essas instituições podem catalogar inúmeros conjuntos de dados (dcat:Dataset1), que podem ter inúmeros recursos (dcat:Distribution) relacionados e as organizações publicados podem cadastrar também inúmeros reúsos (dcat:Dataset2), que podem estar associados a inúmeros conjuntos de dados (dcatDatset1). A estrutura proposta pode ser visualizada na Figura 21:

Figura 21 - Diagrama com classes e propriedades do DCAT-3



Fonte: elaborada pela autora (2024)

A classe dcat:Catalog possui alguns metadados descritivos com informações básicas sobre o Portal, sendo eles: 'Título', 'Descrição', 'Homepage / Link' e 'Lista de Conjuntos de Dados'. Para representá-los, foram utilizadas propriedades advindas da classes dcat:Resource '*title*' e '*description*' (originárias do DC) e as propriedades específicas da classe dcat:Catalog '*dataset*' e '*homepage*'.

Para a representação das Organizações que catalogam os dados no Portal, foi utilizada a classe foaf:Organization, como recomendado pelo DCAT-3, advinda do vocabulário FOAF e específica para descrição de instituições. Além de propriedades do FOAF, foram utilizadas propriedades do DC e do VCard, vocabulários interoperáveis entre si. Essa possibilidade de combinar diferentes vocabulários para realizar uma descrição mais precisa demonstra a utilidade da adoção do RDF para a representação de recursos informacionais.

Para correspondência com os metadados que possuem valores controlados, como ‘Tipo de organização’, ‘Natureza jurídica’ e ‘Esfera’, foi utilizada a propriedade do DC ‘*type’*, que requer uma descrição mais detalhada, especificando os valores que podem ser preenchidos em cada campo. Além dessa, foram utilizadas as propriedades do DC ‘*identifier*’, ‘*relation*’ e ‘*description*’. Do FOAf, provieram as propriedades ‘*name*’, ‘*thumbnail*’ ou ‘*logo*’ (para registro da imagem do logo oficial do órgão), ‘*nick*’ (para identificação da sigla) e ‘*homepage*’.

Para o metadado ‘E-mail da organização’, também enxergam-se duas opções: ‘*mbox*’, do Foaf, ou ‘*has E-mail*’, do Vcard, sendo a primeira mais restrita e formal, um identificador único associado a uma organização, e a segunda, mais genérica e flexível. Nesse cenário, considera-se que ambas poderiam satisfazer o propósito, porém acredita-se que, inicialmente, é preferível adotar a opção ‘*hasEmail*’ para um preenchimento mais maleável. Outras propriedades advindas do Vcard, que é um padrão utilizado para representar informações de contato como nomes, endereços, telefones e e-mails, são: ‘*region*’, ‘*locality*’, ‘*street-address*’, ‘*extend-address*’, ‘*postalcode*’, ‘*tel*’, e ‘*hours*’.

O DCAT-3 permite a modelagem de duas classes iguais, porém com propriedades distintas, que foi a abordagem utilizada para representar os conjuntos de dados e reúsos, ambos representados por uma classe dcat:Dataset, porém cada um com um conjunto específico de metadados. Essa possibilidade abrange, inclusive, a parametrização de metadados obrigatórios e

opcionais na implementação de um perfil de aplicação. Entende-se que, inicialmente, uma organização simplificada traria clareza e objetividade na apresentação da proposta. Para a classe dcat:Dataset1, representando os Conjuntos de Dados, a maioria das propriedades são advindas do DC, sendo elas: '*title*', '*publisher*', '*source*', '*description*', '*license*', '*creator*', '*accrualPeriodicity*', '*spatial*', '*valid*', '*contributor*', '*identifier*', '*conformsTo*', '*issued*', '*modified*' e '*temporal*'. Outras propriedades, originárias do DCAT-3, foram: '*contactPoint*', '*themeTaxonomy*', '*keyword*', '*startDate*', '*endDate*', '*endpointURL*' e '*distribution*'. Para o metadado 'Descontinuado', que visa indicar quando um conjunto de dados não sofrerá mais atualizações, não foi localizada uma propriedade satisfatória no DCAT-3, então utilizou-se a propriedade '*versionNotes*' do vocabulário Adms, que contém informações sobre a versão do atributo e é uma extensão do DCAT-3.

Para alguns metadados, entende-se que existe mais de uma opção viável para representação, como 'Conjuntos de dados associados (interoperabilidade)', que pode ser representado pela propriedade do DCAT-3 '*qualifiedRelation*' ou a propriedade do DC '*relation*', caso seja descrito um relacionado complexo ou a propriedade do DCAT-3 '*dataset*', que lista os *datasets* relacionados, em caso de um relacionamento simples. Acredita-se que uma avaliação definitiva seria melhor realizada em uma etapa de implementação real e conceitual, contudo, julgou-se que, em um cenário para a promoção da interoperabilidade, é mais adequado adotar certo nível de formalização para garantir consistência e padronização, adotando-se, assim, a propriedade '*relation*' do DC.

Para a classe dcat:Dataset2, que representa a categoria Reúso, foram utilizadas as propriedades do DCAT-3 '*access URL*', '*dataset*', '*keyword*', '*contactPoint*' e '*startDate*'; as propriedade do DC utilizadas foram '*title*', '*description*', '*type*', '*relation*', '*creator*', '*issued*' e '*modified*'; e uma propriedade do vocabulário FOAF, para abranger o arquivo de imagem que representa o reúso: '*thumbnail*'. No caso da seleção de metadados para representar duas associações presentes na categoria Reúso, 'Organização a qual o reúso está associado' e 'Conjuntos de Dados de origem', embora houvesse algumas opções, como no caso relatado acima, na categoria Conjuntos de Dados, optou-se pelas propriedades '*relation*', do DC, e '*dataset*', do DCAT,-3 para representá-los, respectivamente, pois entende-se que, nesse caso, é um relacionamento simples, de caráter descritivo, em que o primeiro relaciona o Reúso a uma Organização e o segundo lista os Conjuntos de dados utilizados para a criação do reúso.

Por fim, para a representação dos recursos, foi selecionada a classe dcat:Distribution, que objetiva descrever cada versão de um recurso de um conjunto de dados. Para isso, foram utilizadas as propriedades do DCAT-3 '*access URL*', '*startDate*', '*hasCurrentVersion*', '*hasVersion*', '*previouVersion*', '*keyword*' e '*byteSize*'; do DC, utilizaram-se as propriedades '*title*', '*type*', '*description*', '*issued*', '*modified*', '*language*', '*access rights*' e '*conformsTo*'.

Para o metadado 'Formato do recurso', vislumbra-se duas possibilidades de correspondência, que são a propriedade '*format*', do DC, e a propriedade '*mediaType*', do DCAT, sendo a primeira mais genérica, não exigindo formatos padronizados, e a segunda mais formal, devendo se ater ao identificador *Multipurpose Internet Mail Extension* (MIME). A primeira opção simplifica o preenchimento e a segunda oferece a opção de controle de vocabulário, aumentando a padronização e a interoperabilidade. O uso de vocabulário controlado mitiga os erros advindos de preenchimento com a utilização de linguagem natural e, por isso, recomenda-se a utilização da propriedade '*mediaType*', do DCAT-3. Além desses, ainda foi utilizada a propriedade '*versionNotes*', do Adms, para fornecer informações sobre atualizações ou alterações na versão.

Certas propriedades são repetitivas e podem ser utilizadas em várias classes, e algumas provêm exclusivamente de outros vocabulários. Para todos os metadados de 'Título', em todas as classes, por exemplo, foi utilizada a propriedade '*title*', advinda do DC.

Durante a validação realizada, o próprio DCAT-3 indica sugestões de qual vocabulário utilizar para encontrar uma propriedade que ele reutiliza, como a utilização das propriedades '*title*' e '*description*', do DC, e da classe '*Organization*', do FOAF, demonstrando seu princípio de interoperabilidade e facilitando a troca de informações com outras instituições que o utilizem, ou algum dos padrões relacionados.

Segundo Grácio (2002), a simplicidade na descrição de um recurso é um dos motivos para se utilizar um padrão de metadados, e para Alves (2005), a identificação desses padrões e métodos de representação da informação representa um desafio para a comunidade científica. Acredita-se que o DCAT-3 satisfaz, no contexto estudado, os requisitos para proporcionar uma descrição simples e completa dos dados abertos do Portal e promover a busca e recuperação desses dados de forma eficiente.

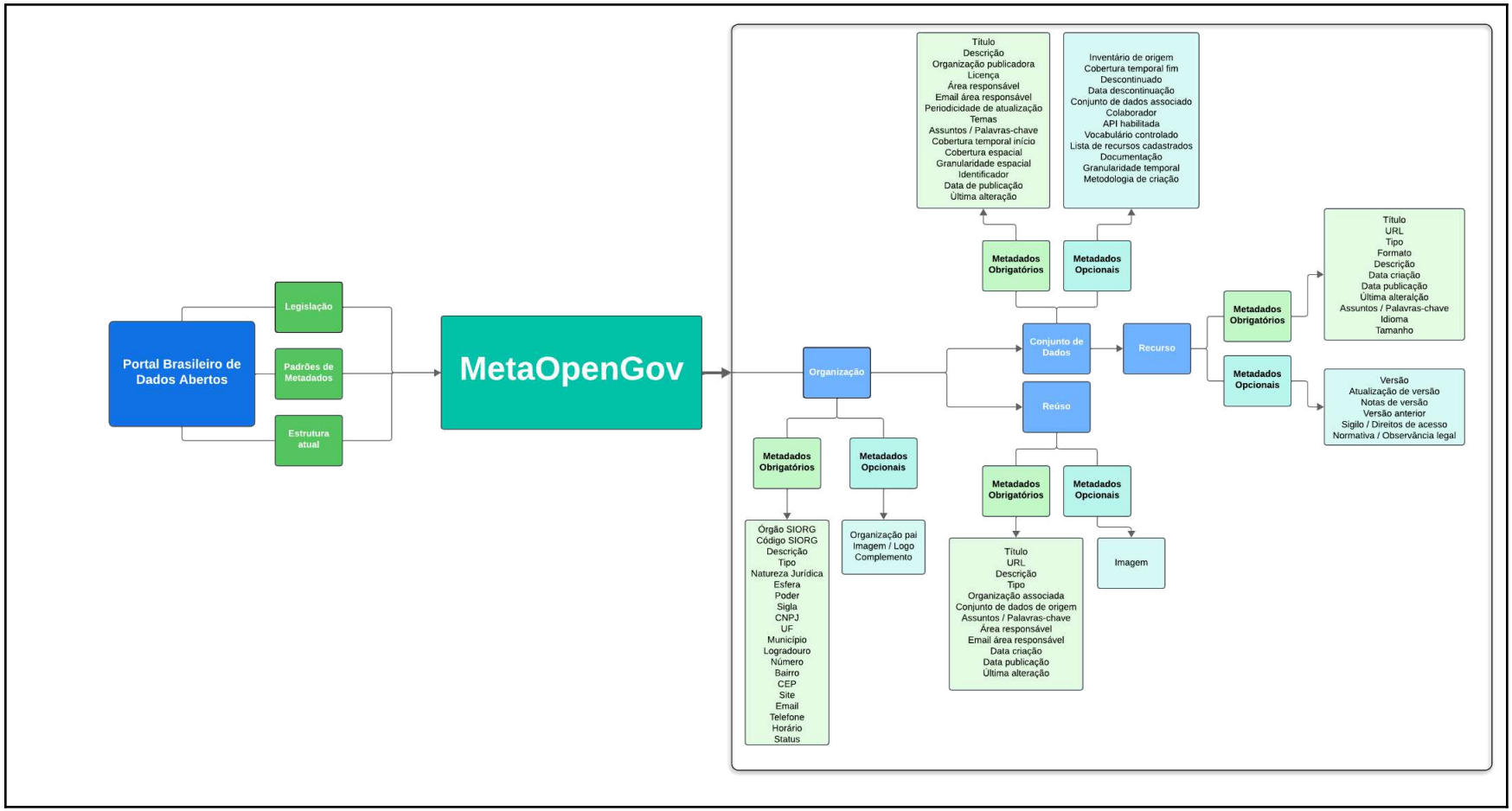
O DCAT-3 é um vocabulário flexível, que permite diversas combinações entre suas classes e propriedades para adequação a cenários específicos, permitindo, inclusive, a adição de

extensões que não sejam contempladas por ele para oferecer uma solução de descrição completa, padronizada e interoperável. Sendo assim, essa proposição é apenas uma opção entre incontáveis possibilidades oferecidas por esse padrão e ressalta-se que o estudo e a atualização constante é requisito fundamental para que a solução proposta seja funcional e pertinente com o domínio utilizado.

6.2 Análise da Proposta

O modelo proposto foi batizado de '**MetaOpenGov**' em um alusão à metadados para os DGA e é apresentado resumidamente na Figura 22:

Figura 22 - Modelo de Metadados para DGA - MetaOpenGov



Fonte: Elaborado pela autora (2024)

Dos 84 metadados sugeridos na proposta, indica-se a correlação de 59 com metadados do e-PMG, como demonstrado no Quadro 22, no Apêndice A, e, consequentemente, aos padrões mapeados por ele. Vê-se que o e-PMG possui informações pertinentes e válidas no contexto atual, porém as suas desobrigatoriedade e desatualização contribuem para sua defasagem e o seu desuso. Quanto aos metadados da Cartilha Técnica para Publicação de Dados Abertos, contemplou-se todos na proposta, pois entende-se que a indicação desses documentos como padrão para descrição deve ser atendida para que essa proposta seja consonante com as diretivas governamentais.

Em relação às Boas práticas para dados na web (Lóscio; Burle; Calegari, 2019), todos os metadados sugeridos puderam ser correlacionados com alguma delas, o que demonstra a pertinência e a relevância da seleção desses metadados, como demonstrado no Quadro 21:

Quadro 21 - Boas práticas para dados na web relacionados a metadados da proposta

Categoria	Boas práticas para dados na web	Metadados da Proposta	Nível de descrição
Metadados	Boa prática para dados na web nº 1 - Fornecer metadados	Todos os metadados da proposta	Organização Conjuntos de dados Recursos Reúso
	Boa prática para dados na web nº 2 - Fornecer metadados descritivos	Tipo de organização Natureza jurídica Esfera Órgão SIORG Código SIORG Poder Organização Pai Imagem / Logo do órgão Sigla CNPJ UF Município Logradouro Número Bairro Complemento CEP Site da organização E-mail da organização	Organização

		Título Organização publicadora Descrição Área técnica responsável E-mail da área técnica responsável Temas Assuntos / Palavras-chave Cobertura temporal início Cobertura temporal fim Granularidade espacial Descontinuado Data de descontinuação Colaborador Data de publicação Última alteração Documentação Granularidade temporal Metodologia de criação	Conjunto de dados
		Título Descrição Data de criação Data de publicação Última alteração Assuntos / Palavras-chave Idioma Sigilo / Direitos de acesso Normativa / Observância legal	Recurso
		Título Descrição Imagem Assuntos / Palavras-chave Área técnica responsável E-mail da área técnica	Reúso
	Boa prática para dados na web nº 3 - Fornecer metadados estruturais	Item de inventário de origem Conjunto de dados associados Lista dos recursos cadastrados	Conjunto de dados

		Tipo Formato Tamanho	Recurso
		Tipo Organização a qual o reúso está associado Conjunto de dados de origem	Reúso
Licença de dados	Boa prática para dados na web nº 4 - Fornecer informações sobre a licença de dados	Licença de uso	Conjunto de dados
Procedência de dados	Boa prática para dados na web nº 5 - Fornecer informações de procedência de dados	Organização publicadora Item de inventário de origem Área técnica responsável Conjuntos de dados associados Colaborador	Conjunto de dados
		URL	Recurso
		URL Organização à qual o reúso está associado Conjunto de dados de origem Área técnica responsável	Reúso
Versionamento de Dados	Boa prática para dados na web nº 7 - Fornecer indicador de versão	Versão	Recurso
	Boa prática para dados na web nº 8 - Fornecer o histórico de versão	Atualização de versão Notas de versão Versão anterior	Recurso
Identificadores de dados	Boa prática para dados na web nº 9 - Usar URIs persistentes como identificadores de conjuntos de dados	Identificador	Conjunto de dados
Formatos de dados	Boa prática para dados na web nº 12 - Usar formatos de dados padronizados legíveis por máquinas	Formato	Recurso
	Boa prática para dados na web nº 13 - Usar representações de dados que sejam independentes de localidade (locale neutral)	Cobertura espacial Granularidade espacial	Conjunto de dados
	Boa prática para dados na web nº 14 - Fornecer dados em formatos múltiplos	Formato	Recurso

Vocabulários de dados	Boa Prática para dados na web nº 15 - Reutilizar vocabulários, dando preferência aos padronizados	Vocabulário controlado associado	Conjunto de dados
Acesso a dados	Boa Prática para dados na web nº 20 - Fornecer acesso em tempo real	URL	Recurso
		URL	Reúso
	Boa Prática para dados na web nº 21 - Fornecer dados atualizados	Periodicidade de atualização Data de publicação Última alteração	Conjunto de dados
		Data de criação Data de publicação Última alteração	Recurso
		Data de criação Data de publicação Última alteração	Reúso
	Boa Prática para dados na web nº 22 - Fornecer uma explicação para os dados que não estão disponíveis	Descontinuado Data de descontinuação	Conjunto de dados
APIs de acesso a dados	Boa Prática para dados na web nº 23 - Disponibilizar dados por meio de uma API	API habilitada	Conjunto de dados
Feedback	Boa Prática para dados na web nº 29 - Coletar feedback de consumidores de dados	E-mail Telefone	Organização
		E-mail da área técnica	Conjunto de dados
		E-mail da área técnica	Reúso
Enriquecimento de dados	Boa Prática para dados na web nº 31- Enriquecer dados por meio da geração de novos dados	Conjuntos de dados associados	Conjunto de dados
		Conjunto de dados de origem	Reúso
Republicação de dados	Boa Prática para dados na web nº 35 - Citar a publicação original do conjunto de dados	API habilitada	Conjunto de dados
		Conjuntos de dados de origem	Reúso

Fonte: elaborado pela autora (2024).

O desenvolvimento de sistemas e aplicações geralmente perpassa uma equipe de profissionais de áreas diversas, mas nem sempre se prevê um profissional da informação, que tem um olhar apurado e cuidadoso para questões de gestão e organização da informação que impactam todos os processos do ciclo de vida dos dados, incluindo sua representação e recuperação, que são refletidos diretamente na experiência do usuário. Pretendeu-se recomendar aprimoramentos na representação descritiva de dados abertos do Portal Brasileiro de Dados Abertos, uma importante ferramenta de disponibilização de informações públicas, sob a perspectiva da CI e seus instrumentos, e espera-se impactar positivamente o usuário consumidor, geralmente não especialista na manipulação de dados.

Essa proposição foi realizada com fundamentação nos conceitos da OI, com foco principalmente na representação da informação por meio do uso de metadados, com o objetivo final de proporcionar uma melhor compreensão dos dados apresentados e uma recuperação eficiente e satisfatória para o usuário. Conforme destacado na fundamentação teórica, embora haja distinções entre a representação descritiva e a representação temática, ambas são complementares e interdependentes. Por essa razão, a proposta também incluiu a temática do controle de vocabulário, reconhecendo sua relevância na padronização e recuperação de dados.

Como evidenciado por Arakaki (2019), cada tipo de metadado desempenha um papel específico na representação de um recurso informacional. Nesse sentido, buscou-se incluir elementos diversificados, capazes de oferecer uma descrição rica e detalhada, apta a atender às diferentes necessidades informacionais. Espera-se contribuir para a mitigação da heterogeneidade dos dados com o oferecimento de opções padronizadas de descrição que favoreçam a interoperabilidade e a integração entre conjuntos de dados, por meio da observação da semelhança entre campos.

7 CONCLUSÕES

A presente dissertação foi desenvolvida com o objetivo geral de propor um modelo de metadados para a representação e a organização da informação no Portal Brasileiro de Dados Abertos e, para isso, foi realizada uma pesquisa de natureza aplicada, dividida em quatro etapas, cada uma correspondente a um objetivo específico.

Na Etapa 1, realizou-se uma revisão integrativa de literatura para identificar, na literatura brasileira, modelos, métricas, *frameworks* e instrumentos similares que abordem a publicação, a avaliação, o ciclo de vida ou o tratamento de dados abertos ou DGA, a fim de evidenciar abordagens, técnicas e métodos utilizados. Essa revisão analisou os trabalhos selecionados de acordo com o momento do tratamento dos dados, categorizando-os entre os que se destacaram em etapas anteriores à publicação dos dados; os que destacaram dados já publicados ou os que abordaram ambos os momentos; as tipologias dos dados abertos investigados, avaliando se foram abordados dados abertos ou DGA de tipos específicos ou generalizados; os conteúdos abordados e as temáticas relacionadas à CI. Também foi elaborado um mapeamento de trabalhos correlatos à identificação de métodos, abordagens e boas práticas utilizados para abordar os DGA no cenário nacional.

Deparou-se, ainda, com alguns desafios recorrentes, alguns dos quais guiaram a presente pesquisa, como a necessidade crucial de padronização na representação dos DGA para garantir sua recuperação, seu uso e reúso e, principalmente, a interoperabilidade com outros sistemas.

Em relação à Etapa 2, foram mapeadas as principais iniciativas, documentos e legislações de dados abertos do governo federal brasileiro por meio de uma pesquisa documental, que foram analisados sob a perspectiva da CI, durante o segundo semestre de 2024. Nessa fase, foram exploradas a Lei Nº 12527/2011 (LAI), o Decreto Nº 8777/2016, documentos relacionados à INDA, o e-PING, o-PMG, o VCGE e o Painel de Monitoramento de Dados Abertos, entre outros instrumentos da Política de Dados Abertos. Esta etapa foi determinante para identificar o contexto atual

Na Etapa 3, foi realizado um mapeamento do contexto dos dados abertos brasileiros, que identificou e analisou as tipologias e os formatos dos dados abertos publicados, padrões e metadados utilizados no Portal Brasileiro de Dados Abertos, por meio de uma pesquisa documental e da observação direta no Portal Brasileiro de Dados Abertos, instrumento oficial de disponibilização de dados do governo federal. A partir desse mapeamento, foi possível identificar

as tipologias, os formatos de dados abertos publicados, os padrões e os metadados utilizados. Observou-se certa fragmentação e algumas inconsistências na apresentação das informações, além da tímida utilização de padrões.

Por fim, na Etapa 4, foi apresentada uma proposta de estrutura de organização para descrição no Portal, com a divisão em categorias e uma proposta de metadados para os dados abertos descritos no Portal Brasileiro de Dados Abertos. Essa proposta foi composta por 84 metadados, sendo quatro para descrever informações básicas do Portal, 23 referentes à categoria Organização, 27 referentes à categoria Conjuntos de Dados, 17 referentes à categoria Recurso e 13 referentes à categoria Reúso.

Desses 84 metadados, foram sugeridos como obrigatórios 61, considerados imprescindíveis para a identificação e compreensão dos dados, como, por exemplo, título e descrição, os que necessariamente irão existir, como data de publicação e última alteração, e como opcionais, 23 metadados que contêm informações complementares ou que não existem em todos os contextos, como versão anterior ou vocabulário controlado associado. O aumento expressivo na quantidade de metadados em relação à estrutura atual reflete a necessidade de prover o máximo de informações possíveis ao usuário para que ele possa tomar decisões sobre a pertinência do conjunto de dados para satisfazer sua necessidade informacional e também para possibilitar o reúso dos dados em novas aplicações, que requerem o máximo de detalhamento possível.

Além da proposição conceitual, fundamentada na legislação e em boas práticas identificadas na literatura científica, propôs-se também a correlação dos campos com propriedades e classes do padrão DCAT-3, o que viabiliza a implementação da proposta. A partir desse panorama, fazem-se algumas considerações sobre os resultados obtidos que são convergentes à literatura, o que demonstra uma homogeneidade nas práticas e abordagens sobre os DGA, mas também acarreta preocupações sobre a recorrência de desafios ao decorrer do tempo.

No que concerne à organização da informação, uma conclusão evidente é que não basta disponibilizar os dados por força da lei, é necessário que esses dados estejam acessíveis de forma organizada e compreensível para o usuário. Para isso, é muito proveitoso que o profissional da informação esteja envolvido no desenvolvimento das políticas de dados abertos.

Algumas recomendações práticas que podem beneficiar diretamente a representação da informação no Portal Brasileiro de Dados Abertos são a utilização de vocabulários controlados atualizados. Ainda que, devido à diversidade dos órgãos e das tipologias dos DGA, não seja viável a utilização de um único vocabulário geral para descrição, como o VCGE, a adoção de vocabulários dentro de cada área específica norteia a compreensão dos dados e facilita o intercâmbio entre dados semelhantes.

7.1 Contribuições da dissertação

As contribuições desta dissertação são tanto teóricas quanto práticas, na medida em que apresentaram-se discussões pertinentes e pautadas na literatura científica.

Como contribuição inicial desta pesquisa, tem-se a centralização dos documentos e sua descrição. Apesar da dificuldade em identificar e selecionar os materiais para consulta, eles foram essenciais para o desenvolvimento da proposta e, por isso, nota-se a escassez de publicações que guiem o usuário consumidor pelo Portal Brasilerio de Dados Abertos e suas funcionalidades. Assim, primeiramente, desenvolveu-se um mapeamento de abordagens de representação e recuperação dos DGA na literatura nacional e internacional (Oliveira; Nascimento Silva, 2024) e, a seguir, mostraram-se os modelos, as métricas e os *frameworks* desenvolvidos para tratar os DGA no cenário nacional (Oliveira; Nascimento Silva, 2025 no prelo[*aguardando publicação]*).

Esses estudos teóricos culminaram em um mapeamento das iniciativas de dados abertos no Brasil, e fundamentaram a proposta de estrutura de metadados para a representação de dados abertos no Portal Brasileiro de Dados Abertos, com a utilização do DCAT-3, e de padrões interoperáveis com esse padrão para a validação da proposta.

Como produtos desta dissertação, foram gerados uma proposta com recomendações para organização da informação e de representação de dados abertos com metadados, que culminou em, um modelo de metadados intitulado: MetaOpenGov. Este produto promove impacto social ao permitir que cidadãos utilizem os dados abertos do governo de forma mais simplificada e organizada, com informações disponibilizadas de forma padronizada, com detalhamento, subsidiando a governança dos dados de diversas temáticas, e sua correspondência com um padrão de metadados reconhecido e recomendado internacionalmente.Além disso, os produtos abrangem discussões sobre a gestão e a tecnologia, linha na qual essa dissertação está enquadrada, pautando

questões de interoperabilidade, reúso e padrões de descrição para recursos informacionais em ambiente digital.

Destaca-se que o modelo MetaOpenGov será publicado em 2025, no Observatório de Dados Abertos: https://observatoriodedadosabertos.eci.ufmg.br/, site do projeto de pesquisa em que a dissertação está vinculada, sendo mais um produto acadêmico para organização, recuperação e reúso dos dados abertos do governo brasileiro.

Espera-se que esse trabalho represente uma base para discussões sobre organização e representação da informação no cenário de dados abertos nacional, fortalecendo a ligação entre os DGA e a CI, e inspire o desenvolvimento de novos modelos e proposições que abranjam o profissional da informação e a interação de diversos atores.

7.2 Desafios e limitações

Um desafio encontrado no desenvolvimento deste trabalho foi a fragmentação, dispersão e desatualização de publicações orientadoras, como normas, diretivas, legislações e manuais, além de *links* inexistentes ou inacessíveis. Uma possível solução seria centralizar a concentração dessa documentação em um repositório, assim tornando possível mapear esses registros e, até mesmo, obter um panorama do desenvolvimento e evolução dessa documentação.

Houve limitações que restringiram o alcance do estudo. Inicialmente, uma questão mencionada anteriormente, tem-se a dependência do poder público para uma implementação real da proposta, o que representa, também, um impedimento da validação com dados reais e, ainda que isso fosse concretizado, a falta de obrigatoriedade de utilização das soluções e ferramentas propostas levará, inevitavelmente, à obsolescência, restringindo sua relevância ao longo do tempo.

Além disso, os dados obtidos por meio de observação direta podem ser imprecisos e estão sujeitos à subjetividade da pesquisadora, como qualquer estudo dessa natureza. Outro entrave dessa perspectiva foi o acesso limitado aos metadados utilizados no sistema, estando acessíveis apenas aqueles disponibilizados para os usuários e os apresentados nos manuais para publicadores de dados, o que pode acarretar em uma análise limitada.

Por fim, a projeção de pesquisa e de testes com usuários representa outra limitação, pois impossibilita a identificação de ajustes para atender adequadamente o público-alvo.

Essas limitações indicam oportunidades para estudos futuros que abordem diretamente as necessidades dos usuários, validem as propostas em contextos reais e explorem estratégias para garantir uma implementação sustentável e contínua.

7.3 Trabalhos futuros

Com base nas contribuições e limitações apresentadas, sugerem-se algumas possibilidades para o desenvolvimento de trabalhos futuros, como realizar uma validação prática com um conjunto de dados genuíno, possibilitando avaliar a aplicabilidade e a eficácia das propostas em um contexto mais próximo da realidade para identificar eventuais ajustes necessários. Além disso, o desenvolvimento de um protótipo de *design*, mais aprofundado, voltado para a organização e apresentação visual da estrutura seria uma complementação atrativa para a proposta apresentada. Esse protótipo possibilitará uma análise mais detalhada da usabilidade e da clareza na transmissão da informação, contribuindo para aprimorar a forma como os dados são apresentados e interpretados.

Essas ações abrem caminho para uma análise mais prática e orientada ao usuário, reforçando o potencial de impacto social dos resultados na área. A realização dessa pesquisa com o usuário pode ser um adendo valioso para verificar a adequação da proposta às necessidades informacionais do cidadão e detectar pontos de melhoria, visando aumentar a participação cidadã.

7.4 Considerações finais

Os profissionais da Biblioteconomia e da Ciência da Informação tem expertise em métodos de classificação e organização da informação, além de dominarem análises e instrumentos que regem os demais fluxos informacionais. Ainda assim, esses profissionais muitas vezes são negligenciados ao não comporem equipes que se debruçam sobre a criação, avaliação e manutenção de metadados.

A inclusão desses profissionais é fundamental devido à sua capacidade e excelência para trabalhar com representação descritiva e temática, o que os torna aptos a contribuir significativamente para a proposição e implementação de metadados.

A tecnologia, por si só, não é suficiente para resolver os desafios da gestão da informação, como afirma a atemporal observação de Davenport. É essencial pensar além das ferramentas tecnológicas e considerar fatores como o usuário e sua interação com o fluxo informacional. A proposta apresentada é fundamentada na organização da informação, refletindo uma abordagem que incorpora não apenas tecnologias, mas também princípios e métodos da Ciência da Informação, complementados por instrumentos da Biblioteconomia, potencializando a eficiência na organização da informação.

A representação da informação em ambientes digitais é uma área de grande potencial profissional para os bibliotecários e cientistas da informação. A discussão, criação e avaliação de metadados está no rol de competências detidas por esses profissionais e a solução apresentada nesta dissertação é viável e real, inclusive com a disponibilidade e preparação da autora para atuar na implementação e gestão desse modelo.

Em síntese, a inserção de bibliotecários em equipes de desenvolvimento de produtos e soluções digitais não é apenas viável, mas também essencial. Com a combinação de suas habilidades com as tecnologias emergentes, é possível criar soluções mais eficientes e acessíveis, fortalecendo a organização da informação em diversos contextos.

REFERÊNCIAS

AGENZIA PER I'ITALIA DIGITALE. **DCAT-AP_IT v1.1 - Profilo italiano di DCAT-AP**. Disponível em: https://www.dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0. Acesso em: 30 nov. 2024.

AGUNE, R. M.; GREGORIO FILHO, A. S.; BOLLIGER, S. P. Governo aberto SP: disponibilização de bases de dados e informações em formato aberto. In: **CONGRESSO CONSAD DE GESTÃO PÚBLICA**, 3., 2010. Anais eletrônicos... Brasília, 2010. Disponível em: https://bit.ly/42EO6As. Acesso em: 10 fev. 2024.

ALBANO, C. S. **Dados governamentais abertos**: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor. 2014. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2014. Disponível em: https://teses.usp.br/teses/disponiveis/12/12139/tde-03062014-170642/pt-br.php. Acesso em: 29 nov. 2024.

ALBERTONI, R *et al* (ed.). **Data Catalog Vocabulary (DCAT) - Version 3**. 2024. Disponível em: https://www.w3.org/TR/vocab-dcat-3/. Acesso em: 15 maio 2024.

ALCARAZ MARTÍNEZ, R. **Dublin Core**: guía de uso. Universitat de Barcelona. 2023. 20 p. Disponível em: https://diposit.ub.edu/dspace/bitstream/2445/203374/1/dublin_core_guia_uso_ruben_alcaraz.pdf. Acesso em: 30 nov. 2024.

ALMEIDA, J. P. D. ***A framework for exploiting open data to improve spatial keyword query applications***. 2021. 167 f. Tese (Doutorado) - Curso de Ciência da Computação, Universidade Federal da Bahia, Salvador, 2021. Disponível em: https://repositorio.ufba.br/handle/ri/33699. Acesso em: 15 abr. 2024.

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, *[S. l.]*, v. 8, n. 15, p. 18–40, 2003. DOI: 10.5007/1518-2924.2003v8n15p18. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2003v8n15p18. Acesso em: 30 nov. 2024.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/e1fd7853-781a-46bd-beee-037375fb178f/content. 2010. Acesso em: 10 fev. 2024.

ALVES, R. C. V. **WEB SEMÂNTICA**: uma análise focada no uso de metadados. 2005. 180 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/d05857ef-4085-49c1-8bae-cf082b02701a/content. Acesso em: 30 nov. 2024.

ANDRADE, T. S. **Desafios e possibilidades para o acesso às informações públicas e a transparência ativa na Universidade Federal de Juiz de Fora - MG**. 2020. 199 f. Dissertação (Mestrado) - Curso de Gestão e Avaliação da Educação Pública, Universidade Federal de Juiz de Fora, Juiz de Fora, 2020. Disponível em: https://repositorio.ufjf.br/jspui/handle/ufjf/12204. Acesso em: 16 abr. 2024.

ARAKAKI, F. A.; ALVES, R. C. V.; SANTOS, P. L. V. A. da C. Dublin Core: state of art (1995 to 2015). **Informação & Sociedade**, *[S. l.]*, v. 28, n. 2, 2018. Disponível em: https://periodicos.ufpb.br/index.php/ies/article/view/38012. Acesso em: 30 nov. 2024.

ARAKAKI, F. A. **Padrão de metadados Dublin Core**. 2021. Repositório - FEBAB. Disponível em: http://repositorio.febab.org.br/items/show/6244. Acesso em: 19 set. 2024.

ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família prov. 2019. 139 f. Tese (Doutorado) - Curso de Programa de Pós- Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, 2019. Disponível em: https://repositorio.unesp.br/items/6b7dc0eb-1feb-4489-a81c-bc41ccbb4fb5. Acesso em: 30 nov. 2024.

ARAÚJO, I. P. **Modelo de Apoio à Decisão para Avaliar Dados Governamentais Abertos do Setor Elétrico**. 2022. 154 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade de Brasília, Brasília, 2022. Disponível em: http://icts.unb.br/jspui/bitstream/10482/45118/1/2022_IngridPalmaAra%C3%BAjo.pdf. Acesso em: 16 abr. 2024.

ARAÚJO JÚNIOR, R. H. de; SOUSA, R. T. B. de. Estudo do ecossistema de *Big Data* para conciliação das demandas de acesso, por meio da representação e organização da informação. **Ciência da Informação**, *[S. l.]*, v. 45, n. 3, 2018. DOI: 10.18225/ci.inf.v45i3.4057. Disponível em: https://revista.ibict.br/ciinf/article/view/4057. Acesso em: 7 abr. 2024.

ARRUDA, W. S. **Dados abertos governamentais**: uma proposta de classificação e estruturação para abertura dos dados de IFES. 2019. 133 f. Dissertação (Programa de Pós-Graduação em Administração Pública (PROFIAP)) - Universidade Federal Rural de Pernambuco, Recife. Disponível em: http://tede2.ufrpe.br:8080/tede/handle/tede2/8238. Acesso em: 15 abr. 2024.

ÁVILA, T. J. T. **Uma proposta de modelo de processo para publicação de dados abertos conectados governamentais**. 2015. 220 f. Dissertação (Mestrado) - Curso de Modelagem Computacional do Conhecimento, Universidade Federal de Alagoas, Maceió, 2015. Disponível em: https://www.repositorio.ufal.br/handle/riufal/2209. Acesso em: 16 abr. 2024.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. Bookman Editora, 2013.

BARBIERI, C. **Governança de Dados**: Práticas, conceitos e novos caminhos. Rio de Janeiro: Alta Books, 2019. 288 p.

BARRETO, Cássia Maria. **Modelo de metadados para a descrição de documentos eletrônicos na web.** 1999. 191 f. Dissertação (Mestrado) - Curso de Ciências em Sistemas e Computação, Instituto Militar de Engenharia, Rio de Janeiro, 1999. Disponível em: https://www.repositorio.mar.mil.br/bitstream/ripcmb/844475/1/1999-Cassia.pdf. Acesso em: 14 maio 2024.

BELUZO, J. R. **IDEO Integrador de dados da Execução Orçamentária Brasileira**: um estudo de caso da integração de dados das receitas e despesas nas esferas federal, estadual governo de são paulo, e municipal municípios do estado de são paulo. 2015. 126 f. Dissertação (Mestrado) - Curso de Ciências, Sistemas de Informação, Universidade de São Paulo, São Paulo, 2015. Disponível em: https://www.teses.usp.br/teses/disponiveis/100/100131/tde-27112015-151717/pt-br.php. Acesso em: 16 abr. 2024.

BERBERIAN, C. F. Q.; MELLO, P. J. S. M.; CAMARGO, R. M. P. Governo Aberto: a Tecnologia Contribuindo Para Maior Aproximação entre o Estado e a Sociedade. **Revista do TCU**, n. 131, p. 30–39, 2014. Disponível em: http://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/60. Acesso em: 30 nov. 2024.

BISI, P. H. S. **Recuperação de dados governamentais**: uma análise de aceitação de tecnologia no acesso a dados em planilhas eletrônicas. 2019. 101 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Universidade Estadual Paulista "Júlio de Mesquita Filho", Marília, 2019. Disponível em: https://repositorio.unesp.br/items/3f5386cb-c880-4653-ae75-cdba923d5c71. Acesso em: 04 maio 2024.

BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.3-5, 1968.

BRÄSCHER, M. CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 9., 2008. São Paulo. Anais...São Paulo: USP, 2008. Disponível em: https://brapci.inf.br/index.php/res/v/176535. Acesso em: 10 fev. 2024.

BRASIL. **Central de paineis de dados abertos**. [*s.d.*]a. Disponível em: https://centralpaineis.cgu.gov.br/visualizar/dadosabertos. Acesso em: 28 jul. 2024.

BRASIL. COMITÊ GESTOR DA INDA (CGINDA). **Plano de Ação da INDA 2023-2025**. 2023. Disponível em: https://dados.gov.br/dados/conteudo/plano-de-acao-da-inda. Acesso em: 30 nov. 2024.

BRASIL. **Constituição (1988)**. Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal: Centro Gráfico, 1988.

BRASIL. **Decreto Nº 6.666 de 27 de novembro de 2008**. Institui, no âmbito do Poder Executivo federal, a Infra-Estrutura Nacional de Dados Espaciais - INDE, e dá outras providências. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6666.htm. Acesso em: 30 nov. 2024.

BRASIL. **Decreto Nº 7.724 de 16 de maio de 2012**. Regulamenta a Lei nº 12.527, de 18 de novembro de 2011, que dispõe sobre o acesso a informações previsto no inciso XXXIII do caput do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/decreto/d7724.htm. Acesso em: 30 nov. 2024.

BRASIL. Decreto nº 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal.**Diário Oficial da União,** Atos do Poder Executivo, Brasília, DF, 12 mai. 2016. Seção 1, p. 21-22. 2016c. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 12 maio. 2024.

BRASIL. Decreto nº 9.903, de 8 de julho de 2019. Altera o Decreto nº 8.777, de 11 de maio de 2016, que institui a Política de Dados Abertos do Poder Executivo federal, para dispor sobre a gestão e os direitos de uso de dados abertos. **Diário Oficial da União**, Atos do Poder Executivo, Brasília, DF, 9 jul. 2019. Seção 1, p. 7. Disponível em: https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=09/07/2019&jornal=515&pagina=7&totalArquivos=67. Acesso em: 30 nov. 2024.

BRASIL. **Decreto Nº 10.046 de 9 de out. de 2019**. Dispõe sobre a governança no compartilhamento de dados no âmbito da administração pública federal e institui o Cadastro Base do Cidadão e o Comitê Central de Governança de Dados. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/d10046.htm. Acesso em 30 nov. 2024.

BRASIL. **Decreto Nº 11.529 de 16 de maio 2023**. Institui o Sistema de Integridade, Transparência e Acesso à Informação da Administração Pública Federal e a Política de Transparência e Acesso à Informação da Administração Pública Federal. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/D11529.htm. Acesso em: 30 nov. 2024.

BRASIL. Governo Brasileiro. Comitê Executivo de Governo Eletrônico. **Padrão de Metadados do Governo Eletrônico (e-PMG)**. 2014. 41 p. Disponível em: https://www.gov.br/governodigital/pt-br/governanca-de-dados/PMGVersao1_1.pdf. Acesso em: 10 maio 2024.

BRASIL. Governo Federal. **Conjuntos de dados**. [*s.d.*]b. Disponível em: https://dados.gov.br/dados/conjuntos-dados. Acesso em: 30 nov. 2024.

BRASIL. Governo Federal. Dados abertos. **Bem vindo ao Portal Brasileiro de Dados Abertos**. [*s.d.*]c. Disponível em: https://dados.gov.br/home. Acesso em: 28 jul. 2024.

BRASIL. Governo Federal. Dados abertos. **Saiba como publicar um conjunto de dados**. [*s.d.*]d. Disponível em: https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-conjunto-de-dados. Acesso em: 30 nov. 2024.

BRASIL. Governo Federal. Dados abertos. **Saiba como publicar um reúso**. [*s.d.*]e. Disponível em: https://dados.gov.br/dados/conteudo/saiba-como-publicar-um-reuso. Acesso em: 30 nov. 2024.

BRASIL. Governo Federal. Dados Abertos. **Sobre**. [*s.d.*]f. Disponível em: https://dados.gov.br/dados/conteudo/sobre. Acesso em: 30 nov. 2024.

BRASIL. **Instrução Normativa Nº 4 de 12 de abril de 2012**. Institui a Infraestrutura Nacional de Dados Abertos – INDA. Disponível em: https://www.gov.br/governodigital/pt-br/legislacao/InstrucaoNormativaINDA42012.pdf. Acesso em: 30. nov. 2024.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5o, no inciso II do § 3o do art. 37 e no § 2o do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da República Federativa do Brasil**. Brasília, DF, 18 nov.2011.Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 12 jun. 2024.

BRASIL.Lei Nº 13.709 de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 30 nov. 2024.

BRASIL. **Lei Nº 14.129 de 29 de março de 2021**. Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública e altera a Lei nº 7.116, de 29 de agosto de 1983, a Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), a Lei nº 12.682, de 9 de julho de 2012, e a Lei nº 13.460, de 26 de junho de 2017. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/l14129.htm. Acesso em: 30 nov. 2024.

BRASIL. Ministério da Ciência, Tecnologia e Inovação. Instituto Brasileiro de Informação em Ciência e Tecnologia. **CKAN**: Sistema de Dados de Governo. 2021. Disponível em: https://www.gov.br/ibict/pt-br/assuntos/informacao-para-o-governo/sistema-de-dados-de-governo-ckan. Acesso em: 30 nov. 2024.

BRASIL. Ministério da Economia. Secretaria de Governo Digital. **Cartilha de Governança de Dados do Poder Executivo Federal**. 2022. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/governancadedados/forum-governanca-de-dados/cartilha-de-governanca-de-dados-volume1-8-12.pdf, Acesso em: 30 nov. 2024.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **Governança de Dados**. [*s.d.*]g. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/governancadedados. Acesso em: 30 nov. 2024.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Governo Digital. Dados Abertos. **Sobre Dados Abertos**. [s.d.]h. Disponível em: https://www.gov.br/governodigital/pt-br/dados-abertos. Acesso em: 30 nov. 2024.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **Linha do tempo:** do eletrônico ao digital. 2024. Disponível em: https://www.gov.br/governodigital/pt-br/estrategia-de-governanca-digital/do-eletronico-ao-digital. Acesso em: 30 nov. 2024.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **Infraestrutura Nacional de Dados Abertos**. 2023. Disponível em: https://www.gov.br/governodigital/pt-br/dados-abertos/infraestrutura-nacional-de-dados-abertos. Acesso em: 30 nov. 2024.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **Vocabulário Controlado**. Vocabulário Controlado do Governo Eletrônico - VCGE, interface de comunicação, Interoperabilidade, ePING. 2024. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/registros-de-referencia/vocabulario-controlado-do-governo-eletronico. Acesso em: 30 nov. 2011.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Cartilha Técnica para Publicação de Dados Abertos no Brasil**. 2016. 10 p. Disponível em: https://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Cartilha_Tecnica.pdf. Acesso em: 30 nov. 2024.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **Vocabulário Controlado de Governo Eletrônico (VCGE)**. Ministério do Planejamento, Orçamento e Gestão, Secretaria de Logística e Tecnologia da Informação - Brasília : MP, SLTI, 2016. 63 p. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/VCGE_2_1_0.pdf. Acesso em: 30 nov. 2024.

BRASIL. **Portaria Nº 5 de 14 jul. 2005**. Institucionaliza os Padrões de Interoperabilidade de Governo Eletrônico - e-PING, no âmbito do Sistema de Administração dos Recursos de Informação e Informática – SISP, cria sua Coordenação, definindo a competência de seus integrantes e a forma de atualização das versões do Documento. Disponível em: https://www.gov.br/governodigital/pt-br/legislacao/Portaria_ePING_14_07_2005.pdf. Acesso em: 30 nov. 2024.

BRASIL. **Portaria Nº 92 de 24 de dez. de 2014**. Institui a arquitetura ePING (Padrões de Interoperabilidade de Governo Eletrônico), que define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) na interoperabilidade de serviços de Governo Eletrônico. Disponível em: https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=1&pagina=50&data=26/12/2014. Acesso em: 30 nov. 2024.

BRASIL. **Resolução Nº 3 de 13 de out. de 2017**. Aprova as normas sobre elaboração e publicação de Planos de Dados Abertos, conforme disposto no Decreto nº 8.777, de 11 de maio de 2016. Disponível em: https://repositorio.cgu.gov.br/handle/1/64747. Acesso em: 30 nov. 2024.

BRASIL. Secretaria de Logística e Tecnologia da Informação. **Instrução Normativa Nº 4, de 12 de abril de 2012**. Institui a Infraestrutura Nacional de Dados Abertos - INDA. Diário Oficial da União. Disponível em: https://www.gov.br/governodigital/pt-br/legislacao/InstrucaoNormativaINDA42012.pdf. Acesso em: 30 nov. 2024.

BRASIL. Portal do Governo Brasileiro. Programa de Governo Eletrônico Brasileiro. **Padrões de Interoperabilidade de Governo Eletrônico - e-PING**. 2018. Disponível em: https://eping.governoeletronico.gov.br/. Acesso em: 30 nov. 2024.

BRICKLEY, D.; MILLER, L. **FOAF Vocabulary Specification 0.99**. 2014. Disponível em: http://xmlns.com/foaf/spec/. Acesso em: 30 nov. 2024.

BRITO, J. C. B.; MARTINS, D. L. Revisão sistemática da literatura na Ciência da Informação: uma descrição detalhada dos passos metodológicos. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, Brasil, v. 14, n. 2, p. 24–47, 2023. DOI: 10.11606/issn.2178-2075.v14i2p24-47. Disponível em: https://www.revistas.usp.br/incid/article/view/209021.. Acesso em: 30 nov. 2024.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science (JASIS)**, v.45, n.5, p.351-360, 1991.

CARVALHO, R. C. **Chatbot aplicado à recuperação de informação:** um modelo orientado a metadados. 2022. 214 f. Tese (Doutorado) - Curso de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (Unesp), Marília, 2022. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/68a4207e-25d6-4c9d-a4bb-0177d817f045/content. Acesso em: 04 maio 2024.

CASAES, J. C. C. **Governança de dados abertos governamentais**: *framework* conceitual para as universidades federais, baseado em uma visão sistêmica. 2019. 220 f. Tese (Doutorado) - Curso de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2019. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/211662/PEGC0573-T.pdf?sequence=-1&isAllowed=y. Acesso em: 16 abr. 2024.

CASTRO, F. F. Catalogação Descritiva: necessidade de revisão em suas bases teórico-conceituais?. **Informação & Informação**, *[S. l.]*, v. 25, n. 3, p. 107–134, 2020. DOI: 10.5433/1981-8920.2020v25n3p107. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41888. Acesso em: 30 nov. 2024.

CATARINO, M. E.; CERVANTES, B. M. N.; ALMEIDA, I. A. de. A representação temática no contexto da web semântica. **Informação & Sociedade**, *[S. l.]*, v. 25, n. 3, p. 105–116, 2015. Disponível em: https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/16242. Acesso em: 27 mar. 2024.

CESARINO, M. A. N.. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, 1985.

CGI.br/NIC.br; CENTRO DE ESTUDOS SOBRE TECNOLOGIA WEB (CEWEB). NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. **Guia de incentivo ao re(uso) de dados abertos**. 2016. Disponível em: https://ceweb.br/guias/incentivo-ao-reuso-de-dados-abertos/. Acesso em: 10 fev. 2024.

CONEGLIAN, C. S.; GONÇALEZ, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O Profissional da Informação na Era do Big Data. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, *[S. l.]*, v. 22, n. 50, p. 128–143, 2017. DOI: 10.5007/1518-2924.2017v22n50p128. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p128. Acesso em: 19 jun. 2024.

CONTROLADORIA-GERAL DA UNIÃO (CGU). **Plano de Dados Abertos 2024-2026**. 2024. Disponível em: https://www.gov.br/cgu/pt-br/acesso-a-informacao/dados-abertos/arquivos/plano-de-dados-abertos-2024.pdf. Acesso em: 30 nov. 2024.

CUNHA, P. **Plano de dados abertos**: estudo multicaso para a priorização de dados a partir da transparência passiva na universidade federal de santa catarina. 2020. 140 f. Dissertação (Mestrado) - Curso de Administração Universitária, Universidade Federal de Santa Catarina, Florianópolis, 2020. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/229044/PPAU0229-D.pdf?sequence=-1&isAllowed=y. Acesso em: 30 mar. 2024.

DAVANZO, L.; MOREIRA, W. O vocabulário controlado como ferramenta do processo de organização e recuperação da informação. 2014. Disponível em: https://brapci.inf.br/index.php/res/v/186303. Acesso em: 10 fev. 2024.

DAVENPORT, T. H. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação**. Futura, 2002.

DEKKERS, M. **Asset Description Metadata Schema (ADMS)**. W3C, 2013. Disponível em: https://www.w3.org/TR/vocab-adms/. Acesso em: 30 nov. 2024.

EAVES, D. **The Three Laws of Open Government Data**. 2009. Disponível em: https://eaves.ca/2009/09/30/three-law-of-open-government-data/. Acesso em: 27 mar. 2024.

ESPÍNDOLA, P. L.; SALM JUNIOR, J. F.; ROSA, F.; JULIANI, J. P. Governança de dados aplicada à ciência da informação: análise de um sistema de dados científicos para a área da saúde. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 16, n. 3, p. 274–298, 2018. DOI: 10.20396/rdbci.v16i3.8651080. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8651080. Acesso em: 30 nov. 2024.

EUROPEAN COMMISSION SUPRA-NATIONAL AUTHORITY. **DCAT Application Profile for data portals in Europe**. Disponível em: https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe. Acesso em: 30 nov; 2024.

FANTINI, W. S. **Publicação de dados conectados sobre despesas orçamentárias do governo federal brasileiro**. 2015. 154 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2015. Disponível em: https://repositorio.ufpe.br/handle/123456789/16779. Acesso em: 16 abr. 2024.

FELIX, V. L.; MARTINS, G. K.; BARROS, A. C.; PEREIRA, F. C. M.; LOYOLA, R. A. A Ciência da Informação e sua interdisciplinaridade: analogias e convergências decorrentes da inter-relação com a Biblioteconomia. **Folha de Rosto**, v. 2, n. Especial, p. 5-11, 13 dez. 2016. Disponível em: https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/143/88. Acesso em: 30 nov. 2024.

FERNANDEZ-MOLINA, Juan Carlos. Enfoques objetivo y subjetivo del concepto de información. **Revista Española de Documentación Científica**, v. 17, n. 3, 1994, p. 320-331.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação.** Rio de Janeiro: Editora Ciência Moderna Ltda, 2012.

FERREIRA, S. C. Proposta de metadados descritivos no âmbito do processo de construção de um repositório digital para um bairro educador. **Folha de Rosto**, v. 8, n. 2, p. 59-85, 10 set. 2022. Disponível em: https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/562. Acesso em: 7 abr. 2024.

FINK, A. **Conducting research literature reviews:** From the internet to paper. Sage publications, 2019.

FRAGA, V. S. Brasil. Fundação Escola Nacional de Administração Pública. Diretoria de Desenvolvimento Profissional. **Governança de Dados na Transformação Digital**. Enap, 2022. Disponível em: https://mooc41.escolavirtual.gov.br/course/view.php?id=9131. Acesso em: 30 nov. 2024.

FREITAS, J. L. **Programas de governo aberto na América do Sul**: estudo da participação eletrônica, transparência e dados abertos por meio da teoria da ação comunicativa. 2019. 145 f. Tese (Doutorado) - Curso de Administração e Negócios, Escola de Negócios, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2019. Disponível em: https://tede2.pucrs.br/tede2/handle/tede/8981. Acesso em: 30 mar. 2024.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 1, n. 1, p. 60–90, 2004. DOI: 10.20396/rdbci.v1i1.2089. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089. Acesso em: 1 dez. 2024.

FUJITA, M. S. L. A indexação na catalogação de livros em bibliotecas universitárias: aplicação, educação e futuro. In: FUJITA, Mariângela Spotti Lopes (org.). **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias. São Paulo: Cultura Acadêmica, 2009.

FUJITA, M. S. L.; TOLARE, J. B. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação**, v. 24, n. 2, p. 93-125, 2019. DOI: 10.5433/1981-8920.2019v24n2p93. Acesso em: 10 fev. 2024.

GALVÃO, T. F.; PEREIRA, M. G. Revisões sistemáticas da literatura: passos para sua elaboração. **Epidemiol. Serv. Saúde**, Brasília, v. 23, n. 1, p. 183-184, mar. 2014 . Disponível em: http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742014000100018&lng=pt&nrm=iso. Acesso em: 22 nov. 2024.

GARCIA, P. S. R. **Estudo sobre aplicação de *web* semântica e visualização em dados abertos: proposta de portal para transparência em séries históricas de indicadores sociais, econômicos e financeiros, públicos na *web***. 2011. 155 f. Dissertação (Mestrado) - Curso de Tecnologias da Inteligência e Design Digital, Pontifícia Universidade Católica de São Paulo, São Paulo, 2011. Disponível em: https://tede2.pucsp.br/handle/handle/18088. Acesso em: 30 mar. 2024.

GEIßNER, A. **Modeling institutional research data repositories using the DCAT3 Data Catalog Vocabulary**: a case study on TUDatalib. Institut für Bibliotheks - und Information swissen chaft der Humboldt Universitaet zu Berlin, Germany, 2023. 99 p. Disponível em: https://www.proquest.com/openview/676b80bb2de0c8e3573e8ef90b53fd0e/1?pq-origsite=gscholar&cbl=2026366&diss=y. Acesso em: 30 nov. 2024.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. São Paulo: Atlas, 2008. 200 p.

GIL, A. C. **Como elaborar projetos de pesquisa.** 6 ed.São Paulo: Atlas, 2017.

GOMES, M. S. **Um modelo de avaliação de conformidade de portais de dados abertos de governo**. 2021. 263 f. Dissertação (Mestrado) - Curso de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2021. Disponível em: https://repositorio.ufsc.br/handle/123456789/227173. Acesso em: 16 abr. 2024.

GUEDES, W.; ARAÚJO JÚNIOR, R. H. de. Estudo das similaridades entre a teoria matemática da comunicação e o ciclo documentário. **Informação & Sociedade**, *[S. l.]*, v. 24, n. 2, 2014. Disponível em: https://periodicos.ufpb.br/ojs/index.php/ies/article/view/16498. Acesso em: 14 maio. 2024.

GRÁCIO, J. C. A. **Metadados para a descrição de recursos da Internet**: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. 127 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2002. Disponível em: https://acervodigital.unesp.br/handle/11449/93722. Acesso em: 30 nov. 2024.

HJØRLAND, Birger. What is knowledge organization (KO)?. **KO Knowledge Organization**, v. 35, n. 2-3, p. 86-101, 2008.

IANNELLA, R. et al. **vCard Ontology - for describing People and Organizations**. W3C, 2014. Disponível em: https://www.w3.org/TR/vcard-rdf/. Acesso em: 30 nov. 2024.

IDLab Turtle Validator. Disponível em: http://ttl.summerofcode.be/. Acesso em: 30 nov. 2024.

INSTITUTO DE DADOS ABERTOS. **Guia do Modelo de Maturidade de Dados Abertos**: avaliando a publicação e a utilização de dados abertos. São Paulo: Instituto de Dados Abertos, 2016, 53 p. Disponível em: http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/02/Book-Web-Modelo-de-Maturidade-de-Dados-Abertos.pdf. Acesso em: 30 mar. 2024.

JESUS, A. F. de; SANTAREM SEGUNDO, J. E. . A descrição formal da qualidade de dados publicados na Web: análise do Data Quality Vocabulary (DQV). **Em Questão**, Porto Alegre, v. 29, 2023. DOI: 10.1590/1808-5245.29.129415. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/129415. Acesso em: 11 set. 2024.

KOBASHI, N. Y. Vocabulário controlado: estrutura e utilização. Brasília, Escola Nacional de Administração Pública, 2008. Disponível em: https://repositorio.enap.gov.br/bitstream/1/1289/41/Vocabul%C3%A1rio%20controlado%20-%20estrutura%20e%20utiliza%C3%A7%C3%A3o.pdf. Acesso em: 30 nov. 2024.

KOLLER, S. H.; COUTO, M. C. P. P.; VON HOHENDORFF, J.. **Manual de produção científica**. Penso Editora, 2014.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003. 311 p.

LEMOS, J. M. M. **Um processo para publicação de dados abertos em institutos federais baseado em BPM**. 2017. 147 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2017. Disponível em: https://repositorio.ufpe.br/handle/123456789/25351. Acesso em: 16 abr. 2024.

LIMA, G. Â. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, *[S. l.]*, p. 57–97, 2020. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/22283. Acesso em: 7 abr. 2024.

LIMA, J. L. O.; ALVARES, L.. Organização e representação da informação e do conhecimento. In: LIMA, J. L. O.; ALVARES, L.; CARDOSO FILHO, J. C.; SCHIESSL, M. (org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4, 2012. p. 21-48.

LIRA, Márcio Ângelo Bezerra. **Uma Abordagem Para Enriquecimento Semântico de Metadados Para Publicação de Dados Abertos**. 2014. 95 f. Dissertação (Mestrado em Ciência da 122 Informação) - Centro de Informática, Universidade Federal de Pernambuco, Recife, 2014. Disponível em: https://repositorio.ufpe.br/handle/123456789/11570. Acesso em: 10 nov. 2024.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. (ed). **Boas Práticas para Dados na Web**. Núcleo de Informação e Coordenação do Ponto BR (trad.). W3C, 2019. Disponível em: https://www.w3.org/Translations/DWBP-pt-BR/. Acesso em: 30 nov. 2024.

LÓSCIO, B. BURLE, C. OLIVEIRA, M. CALEGARI, N. Centro de Estudos sobre Tecnologias Web do Núcleo de Informação e Coordenação do Ponto BR (Ceweb.br/NIC.br). **Fundamentos para publicação de dados na web.** São Paulo: Comitê Gestor da Internet no Brasil, 2018. Disponível em: https://ceweb.br/media/docs/publicacoes/1/fundamentos-publicacao-dados-web.pdf. Acesso em: 30 nov. 2011.

LOURENÇO, Cíntia de Azevedo. Representação de informação, metadados, interoperabilidade e recuperação da informação na atualidade. In: II SEMINÁRIO DO GRUPO DE PESQUISA : PESQUISA INTEGRADA EM ORGANIZAÇÃO DO CONHECIMENTO E RECUPERAÇÃO DA INFORMAÇÃO, 2016. **Anais do II seminário do grupo de pesquisa : pesquisa integrada em organização do conhecimento e recuperação da informação.** Belo Horizonte: Mhtx, 2016. Disponível em: https://repositorio.ufmg.br/handle/1843/37904. Acesso em: 14 maio 2024.

LUCID SOFTWARE. **Lucidchart**: Diagramação com inteligência. 2024. Disponível em: https://www.lucidchart.com/pages/pt. Acesso em: 30 nov. 2024.

LUSA, D. A. ***Open smart city view***: uma solução computacional para manipulação e apresentação de dados governamentais. 2016. 82 f. Dissertação (Mestrado) - Curso de Computação, Universidade de Passo Fundo, Passo Fundo, 2016. Disponível em: http://tede.upf.br/jspui/handle/tede/1184#preview-link0. Acesso em: 16 abr. 2024.

MALIN, A. M. B. Interessados e interesses no regime de acesso à informação pública no Brasil. **Coleção Estudos da Informação**, v. 3, p. 1-13, 2013. Disponível em: http://obgi.org/wp-content/uploads/2014/11/Produ%C3%A7%C3%A3o-Intelectual-2013-Interessados-E-Interesses-No-Regime-De-Acesso-%C3%80-Informa%C3%A7%C3%A3o-P%C3%BAblica-No-Brasil.pdf. Acesso em: 04 maio 2024.

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. de F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade**, *[S. l.]*, v. 21, n. 1, 2011. Disponível em: https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/7367. Acesso em: 27 mar. 2024.

MARTINS, L. C. B. **Proposta de arquitetura de publicação automatizada de dados abertos conectados utilizando meta-dados e ontologias**. 2018. 80 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade de Brasília, Brasília, 2019. Disponível em: http://icts.unb.br/jspui/handle/10482/34816. Acesso em: 16 abr. 2024.

MELLO FILHO, L. L.; ARAÚJO JÚNIOR, R. H. Objetos de fronteira: um diálogo entre a ciência da informação e a ciência de dados . **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, *[S. l.]*, v. 26, p. 01–22, 2021. DOI: 10.5007/1518-2924.2021.e77247. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/77247. Acesso em: 14 maio. 2024.

MEY, Eliane Serrão Alves. **Introdução à catalogação**. Brasília: Briquet de Lemos, 1995. 123 p.

MIRANDA, R. C. R. O uso da informação na formulação de ações estratégicas pelas empresas. **Ciência da informação**, v. 28, p. 286-292, 1999. Disponível em: https://www.scielo.br/j/ci/a/r7L9msHr6FfrYpJ5PKk8fsS/. Acesso em: 27 mar. 2024.

MONTEIRO, D. A. A.; SARACENO, R. D. F.. Percepção dos servidores sobre a Lei de Acesso à Informação em uma organização universitária pública: Percepção dos servidores sobre a Lei de Acesso à Informação em uma organização universitária pública. **RP3-Revista de Pesquisa em Políticas Públicas**, n. 2, 2021. Disponível em: https://periodicos.unb.br/index.php/rp3/article/view/37228. Acesso em: 30 nov. 2024.

NASCIMENTO SILVA, P. Acesso à informação no Brasil: política de dados governamentais abertos. *In*: GERALDES, E. *et al* (orgs.). Dez anos da lei de acesso à informação: limites, perspectivas e desafios. São Paulo: INTERCOM, 2022. p. 748.

NASCIMENTO SILVA, P. **Dados governamentais abertos**: métricas e indicadores de reúso. 2018. 322 f. Tese (Doutorado) - Curso de Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2018. Disponível em: https://repositorio.ufmg.br/handle/1843/BUBD-AYNG4U. Acesso em: 15 abr. 2024.

NASCIMENTO SILVA, P. Dados governamentais abertos sobre a COVID-19 no Brasil: avaliação do potencial de reúso. **Em Questão**, Porto Alegre, v. 29, p. 124820, 2023. DOI: 10.19132/1808-5245.29.124820. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/124820. Acesso em: 10 fev. 2024.

NASCIMENTO SILVA, P. Recuperação de dados abertos para inovação cívica e ciência cidadã: uma análise do novo portal brasileiro. **AtoZ: novas práticas em informação e conhecimento**, *[S. l.]*, v. 13, p. 1–6, 2024. DOI: 10.5380/atoz.v13i0.89352. Disponível em: https://revistas.ufpr.br/atoz/article/view/89352. Acesso em: 30 nov. 2024.

OLIVEIRA, C. C. **Ontologias como artefatos para a representação de dados abertos de pesquisa**: uma exploração acerca de uma modelagem experimental de dados da arqueologia através do modelo cidoc conceptual reference model. 2023. 174 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023. Disponível em: https://lume.ufrgs.br/handle/10183/262751. Acesso em: 16 abr. 2024.

OLIVEIRA, D. T.; NASCIMENTO SILVA, P.; MAFRA PEREIRA, F. C. Inteligência Artificial para Recuperação de Dados Abertos: reflexões e proposições para a experiência do usuário no Portal Brasileiro de Dados Abertos. **Brazilian Journal of Information Science: research trends**, Marília, SP, v. 18, p. e024016, 2024. DOI: 10.36311/1981-1640.2024.v18.e024016. Disponível em: https://revistas.marilia.unesp.br/index.php/bjis/article/view/15462.. Acesso em: 30 nov. 2024.

OLIVEIRA, D. T.; NASCIMENTO SILVA, P. Modelos e Frameworks para publicação de dados abertos do governo: uma análise da produção acadêmica brasileira (Aguardando publicação).

OLIVEIRA, D. T.; NASCIMENTO SILVA, P. Representação e recuperação de dados governamentais abertos: uma revisão de literatura. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 22, n. 00, p. e024029, 2024. DOI: 10.20396/rdbci.v22i00.8675828. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8675828. Acesso em: 30 nov. 2024.

OPEN KNOWLEDGE BE. **DCAT-BE**. Linking data portals across Belgium. Disponível em: http://dcat.be/. Acesso em: 30 nov. 2024.

OPEN KNOWLEDGE BRASIL. **Índice de Dados Abertos**. [*s.d.*]a. Disponível em: https://ok.org.br/projetos/indice-dados-abertos/. Acesso em: 30 nov. 2024.

OPEN KNOWLEDGE BRASIL. **Quando surgiram os oito princípios dos dados abertos?**. 2015. Disponível em: https://ok.org.br/noticia/quando-surgiram-os-oito-principios-dos-dados-abertos/. Acesso em: 30 nov. 2024.

OPEN KNOWLEDGE BRASIL. **Por que "open"?**. *[s.d.*]b. Disponível em: https://ok.org.br/dados-abertos/. Acesso em: 30 nov. 2024.

OPEN KNOWLEDGE FOUNDATION. **Why open data**?. *[s.d.]*. Disponível em: https://okfn.org/en/library/why-open-data/. Acesso em: 30 nov. 2024.

ORTEGA, C. D. Contexto de desenvolvimento da organização da informação, com enfoque para a catalogação, na Escola de Ciência da Informação da Ufmg. **Perspectivas em Ciência da Informação**, v. 18, n. 2, p. 182-215, 2013. Disponível em:http://hdl.handle.net/20.500.11959/brapci/35682. Acesso em: 10 fev. 2024.

PALMER, M. SWEDISH AGENCY FOR DIGITAL GOVERNMENT. META SOLUTIONS AB. **DCAT-AP-SE version 3.0.1**. Disponível em: https://docs.dataportal.se/dcat/en/. Acesso em: 30 nov. 2024.

PAULO, J. S.. **Dados abertos governamentais**: uma análise aplicada ao ministério público do trabalho. 2018. 83 f. Dissertação (Mestrado) - Curso de Mestrado Profissional em Gestão Pública, Universidade Federal do Espírito Santo, Vitória, 2018. Disponível em: https://repositorio.ufes.br/server/api/core/bitstreams/811af1b4-4fda-4d26-b660-1b45f3c9c436/content. Acesso em: 30 mar. 2024.

PAVÃO, C. G.; DA COSTA, J. B.; FERREIRA, M. K.; HOROWITZ, Z. METADADOS E REPOSITÓRIOS INSTITUCIONAIS: UMA RELAÇÃO INDISSOCIÁVEL PARA A QUALIDADE DA RECUPERAÇÃO E VISIBILIDADE DA INFORMAÇÃO. **PontodeAcesso**, [S. l.], v. 9, n. 3, p. 103–116, 2015. Disponível em: https://periodicos.ufba.br/index.php/revistaici/article/view/15163. Acesso em: 30 nov. 2024.

PENTEADO, B. E. **Modelo de infraestrutura para publicação de dados abertos governamentais conectados de qualidade**. 2020. 296 f. Tese (Doutorado) - Curso de Ciências da Computação e Matemática Aplicada, Universidade de São Paulo, São Carlos, 2020. Disponível em: https://www.teses.usp.br/teses/disponiveis/55/55134/tde-14092020-175138/pt-br.php. Acesso em: 16 abr. 2024.

PEREIRA, C. M. **Uma abordagem para a publicação de dados abertos ligados obtidos a partir de bases de dados relacionais**. 2012. 156 f. Tese (Doutorado) - Curso de Engenharia Eletrônica e Computação, Instituto Tecnológico de Aeronáutica, São José dos Campos, 2012. Disponível em: http://www.bdita.bibl.ita.br/. Acesso em: 16 abr. 2024.

PEREIRA, D. V. **MODELAGEM E REPRESENTAÇÃO SEMÂNTICA DE DADOS GOVERNAMENTAIS ABERTOS DA PREVIDÊNCIA SOCIAL BRASILEIRA**. 2014. 168 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Universidade Federal Fluminense, Niterói, 2014. Disponível em: https://app.uff.br/riuff;/handle/1/10686. Acesso em: 16 abr. 2024.

PEREIRA, L. M. F. **OGDPub**: uma ontologia para publicação de dados abertos governamentais. 2017. 185 f. Dissertação (Mestrado) - Curso de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2017. Disponível em: https://repositorio.ufsc.br/xmlui/handle/123456789/179907. Acesso em: 16 abr. 2024.

PINHO, M. D. C. Dados Abertos Governamentais: Usuários e apropriações sociais no Brasil. **Boletim de Análise Político-institucional**, n. 25, 2021. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/10494/6/bapi_25_dados_abertos.pdf. Acesso em: 21 fev. 2024.

PINHO, M. D. C.; SILVA, M. P. A. Governo aberto e dados abertos governamentais: um mapeamento e sistematização da produção acadêmica. **Comunicação & Inovação**, v. 20, n. 43, 2019. Disponível em: https://seer.uscs.edu.br/index.php/revista_comunicacao_inovacao/article/view/5711/0. Acesso em: 30 mar. 2024.

PINTO, J. A. ***Government institutions and their documents***: *an ontological approach*. 2020. 2020 f. Tese (Doutorado) - Curso de Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2020. Disponível em: https://repositorio.ufmg.br/handle/1843/34427. Acesso em: 16 abr. 2024.

PIRES, L. B. **Integração, visualização e análise de informações eleitorais usando bancos de dados analíticos e fontes heterogêneas de grande volume**. 2015. 198 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal do Abc, Santo André, 2015. Disponível em: http://biblioteca.ufabc.edu.br/index.php?codigo_sophia=77321. Acesso em: 16 abr. 2024.

POSSAMAI, A. J. **Dados abertos no governo federal brasileiro**: desafios de transparência e interoperabilidade. 2016. 313 f. Tese (Doutorado) - Curso de Ciência Política, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016. Disponível em: https://lume.ufrgs.br/handle/10183/156363. Acesso em: 30 mar. 2024.

PRECHT, P. A. C. **Do mapeamento de dados abertos à modelagem conceitual**: um data warehouse para análise de informações multidimensionais. 2019. 63 f. Dissertação (Mestrado) - Curso de Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019. Disponível em: https://lume.ufrgs.br/handle/10183/199912. Acesso em: 16 abr. 2024.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico:** métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

RIBEIRO, C. J. S.; ALMEIDA, R. F. de. Dados abertos governamentais (open government data): instrumento para exercício de cidadania pela sociedade. **Encontro Nacional de Pesquisa em Ciência da Informação**, v. 12, p. 2568-2580, 2011. Disponível em: http://arq.3rengtt.com.br/wp-content/uploads/2015/09/EnancibXII-RibeiroAlmeida.pdf. Acesso em: 30 mar. 2024.

RIGO, F. **Desenvolvimento da interface e automatização da coleta de dados na Open Smart City View**. 2018. 60 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade de Passo Fundo, Passo Fundo, 2018. Disponível em: http://tede.upf.br/jspui/bitstream/tede/1480/2/2018FernandaRigo.pdf. Acesso em: 16 abr. 2024.

ROCHA, L. L. **Ontologias como um agente para a interligação de dados abertos em contexto governamental**: características e aplicabilidade. 2016. 137 f. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal Fluminense, Instituto de Arte e Comunicação Social. Niterói, 2016. Disponível em: https://app.uff.br/riuff/handle/1/10773. Acesso em: 16 abr. 2024.

SANTANA, M. T. **Uma proposta de publicação de dados do orçamento público na *Web***. 2013. 101 f. Dissertação (Mestrado) - Curso de Ciências, Sistemas de Informação, Universidade de São Paulo, São Paulo, 2014. Disponível em: https://www.teses.usp.br/teses/disponiveis/100/100131/tde-02022014-221236/pt-br.php. Acesso em: 16 abr. 2024.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, *[S. l.]*, v. 21, n. 2, p. 116–142, 2016. DOI: 10.5433/1981-8920.2016v21n2p116. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27940. Acesso em: 5 abr. 2024.

SANTOS, P. L. V. A. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, *[S. l.]*, v. 19, n. 1, p. 146–163, 2014. DOI: 10.5433/1981-8920.2014v19n1p146. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/15251. Acesso em: 30 nov. 2024.

SARACEVIC, T. Ciência Da Informação: Origem, Evolução e Relações. Perspectivas em Ciência Da Informação, v.1, n.1, p.41-62, Jan/jun. 1996. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/22308. Acesso em: 30 nov. 2024.

SENADO FEDERAL. ***Open Government Partnership - OGP***. 2024. Disponível em: https://www25.senado.leg.br/web/atividade/sinfleg/relatorio-de-gestao/saiba-mais-23. Acesso em: 27 mar. 2024.

SEVERINO, A. J. **Metodologia do trabalho científico**. São Paulo: Cortez, 2013. 274 p.

SETZER, Valdemar W. Dado, informação, conhecimento e competência. **DataGramaZero Revista de Ciência da Informação, n. 0**, v. 28, 1999. Disponível em: https://www.ime.usp.br/~vwsetzer/datagrama.html. Acesso em: 10 fev. 2024.

SILVA, L. C. **Publicação de dados de pesquisa científica**: proposta de estruturação semântica de cadernos abertos de pesquisa frente às dimensões da e-science. 2020. 243 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Estadual Paulista, Marília, 2020. Disponível em: https://repositorio.unesp.br/items/360da1e3-cdd3-4dcd-a9c5-8a9db0e355ab. Acesso em: 16 abr. 2024.

SILVA, R. O. **Uma proposta de processo para implantação de dados abertos em instituições públicas brasileiras**. 2018. 123 f. Tese (Doutorado) - Curso de Sistemas e Computação, Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Natal, 2018. Disponível em: https://repositorio.ufrn.br/jspui/handle/123456789/25279. Acesso em: 16 abr. 2024.

SOUZA, D. F. **Transparência ativa para acesso à informação**: uma proposta de modelo com os temas prioritários em instituições federais de ensino superior. 2018. 124 f. Dissertação (Mestrado) - Curso de Administração Pública, Universidade Federal Rural de Pernambuco, Recife, 2018. Disponível em: http://www.tede2.ufrpe.br:8080/tede2/handle/tede2/7608#preview-link0. Acesso em: 16 abr. 2024.

SOUZA, M. T.; SILVA, M. D.; CARVALHO, R. Revisão integrativa: o que é e como fazer. **Einstein (São Paulo)**, v.8, n.1, p.102-10, 2010. Disponível em: https://doi.org/10.1590/S1679-45082010RW1134. Acesso em: 30 nov. 2024.

SOUZA, R. R.; ALMEIDA, M. B.. Representação do conhecimento: identidade ou esvaziamento da ciência da informação?. In: BORGES, Maria Manuel; CASADO, Elías Sanz (org.). **A ciência da informação criadora de conhecimento**: actas do iv encontro ibérico edibcic 2009. Edibcic, 2009. p. 157-165.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n.2, p. 161-173, 2006. Disponível em: https://www.brapci.inf.br/index.php/res/v/33177. Acesso em: 10 fev. 2024.

SPERONI, R. M. **Modelo de referência para indicadores de inovação regional suportado por dados ligados**. 2016. 233 f. Tese (Doutorado) - Curso de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2016. Disponível em: https://repositorio.ufsc.br/xmlui/handle/123456789/172186. Acesso em: 16 abr. 2024.

TAYLOR, R. S. Professional aspects of information science and technology. **Annual Review of Information Science and Technology**, v. 1, n. 1, p. 15-40, 1966.

TOMOYOSE, K. **O Data Catalog Vocabulary (DCAT) para a publicação de dados de pesquisa nos princípios Linked Data**. 2021. 185 f. Dissertação (Mestrado) - Curso de Mestrado em Ciência da Informação, Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de São Carlos, São Paulo, 2021. Disponível em: Universidade Federal de São Carlos. Acesso em: 30 nov. 2024.

TOSIN, T. M. **Um modelo para integração de informações de bases de dados abertos, com uso de ontologias**. 2016. 107 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2016. Disponível em: http://repositorio.jesuita.org.br/handle/UNISINOS/5244. Acesso em: 16 abr. 2024.

VIEIRA, D. I. **Uma plataforma para disponibilização centralizada de dados abertos governamentais como suporte para aplicações no contexto de cidades inteligentes**. 2016. 91 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de São Carlos, Sorocaba, 2016. Disponível em: https://repositorio.ufscar.br/handle/ufscar/8810. Acesso em: 16 abr. 2024.

VISINTIN, L. **Modelo de maturidade de dados abertos**: uma matriz de referência para organizações. 2021. 211 f. Tese (Doutorado) - Curso de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2021. Disponível em: https://repositorio.ufsc.br/handle/123456789/231168. Acesso em: 16 abr. 2024.

WORLDOMETER. **Worldometer**. Disponível em: https://www.worldometers.info/. Acesso em: 04 maio 2024.

W3C. **Conhecendo o W3C**. [*s.d.*]. Disponível em: https://www.w3c.br/conhecendo-o-w3c. Acesso em: 20 jun. 2024.

W3C. ***Linked Data***. 2023. Disponível em: https://www.w3.org/wiki/LinkedData. Acesso em: 27 mar. 2024.

YAMANE, G. A. da C.; CASTRO, F. F. de. O estudo e a identificação dos padrões de metadados para a representação e a recuperação da imagem digital na perspectiva da web. **Em Questão**, Porto Alegre, v. 24, n. 1, p. 145–173, 2018. DOI: 10.19132/1808-5245241.145-173. Disponível em: https://seer.ufrgs.br/index.php/EmQuestao/article/view/71475. Acesso em: 14 maio. 2024.

ZAFALON, Z. R. **SCAN FOR MARC**: Princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o formato MARC21 bibliográfico. 2012. 169f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista, Marília, 2012. https://repositorio.unesp.br/handle/11449/103386. Acesso em: 10 fev. 2024.

ZANELLA, A. *et al*. *Internet of things for smart cities*. ***IEEE Internet of Things journal***, v. 1, n. 1, p. 22-32, 2014. Disponível em: https://ieeexplore.ieee.org/document/6740844. Acesso em: 30 mar. 2024.

APÊNDICE A - Proposta de metadados para o Portal Brasileiro de Dados Abertos

Para melhor compreensão da proposta, foi utilizado um esquema de cores em que as linhas em fundo **Lilás**, representam os metadados identificados na estrutura atual do Portal na subseção 5.2.2 e que foram mantidos, as linhas em fundo **Verde** caracterizam os metadados adicionados e as linhas em fundo **Azul** representam os metadados que sofreram alguma modificação como alteração de visibilidade ou de obrigatoriedade.

Quadro 22 - Proposta de metadados para o Portal Brasileiro de Dados Abertos

Metadado	Tipos de dados	Descrição	Nível	Tipo	Visível? (Sim/Não)	Correspondência e-PMG	Propriedade	Classe	Superclasse	Padrão	Fundamentação	Observações
Título	**String**	**Nome do catálogo**	**Portal Brasileiro de Dados Abertos**	**Obrigatório**	**Sim**	**Título (Title)**	***dcterms:title***	***dcat:Catalog***	***dcat:Dataset***	**Dublin Core**	**Boas práticas 1, 2, 3 e 5**	
Descrição	**String**	**Descrição do catálogo de dados**		**Obrigatório**	**Sim**	**description.resumo (description.abstract)**	***dcterms:description***	***dcat:Catalog***	***dcat:Dataset***	**Dublin Core**		
Homepage / Link	**URL**	**Link de acesso ao catálogo**		**Obrigatório**	**Sim**	**Localização (Location)**	***dcat:landingPage***	***dcat:Catalog***	***dcat:Dataset***	**Dcat**		
Lista de Conjuntos de Dados	**String**	**Relaciona os Conjuntos de Dados do Catálogo**		**Opcional**	**Sim**	**-**	***dcat:dataset***	***dcat:Catalog***	***dcat:Dataset***	**Dcat**		
Tipo de organização	String	Governamental: órgãos e entidades vinculados ao Governo de qualquer unidade federativa. Não Governamental: organizações de direito privado que atuam na sociedade civil e que não pertencem ao Estado.	Organização / Órgão	Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:type*	*foaf:Organization*	*foaf:Agent*	Dublin Core	LAI Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas. § 1º Na divulgação das informações a que se refere o caput, deverão constar, no mínimo: I - registro das competências e estrutura organizacional,	Valores controlados: "Governamental" ou "Não governamental" Alteração de visibilidade
Natureza Jurídica	**String**	**O poder executivo federal é dividido em administração direta (ministérios) e indireta (autarquias, fundações públicas e empresas estatais)**		**Obrigatório**	**Sim**	Contexto Jurídico-administrativo (Mandate)	*dcterms:type*	***foaf:Organization***	***foaf:Agent***	**Dublin Core**		**Valores controlados: "Administração Direta" ou "Administração Indireta"**

Esfera	String	No Brasil, o sistema político está dividido em três esferas: União (Federal), Estados (Estadual) e Municípios (Municipal), além do Distrito Federal (DF)		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	dcterms:type	*foaf:Organization*	*foaf:Agent*	Dublin Core	endereços e telefones das respectivas unidades e horários de atendimento ao público; Boas práticas 1, 2 e 29	Valores controlados: "Federal" ou "Estadual" ou "Municipal"
Órgão SIORG	String	Cadastro oficial dos órgãos e entidades da administração direta, autárquica e fundacional do Poder Executivo Federal junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:name*	*foaf:Organization*	*foaf:Agent*	Foaf		Alteração de visibilidade
Código SIORG	Numério	Referencial numérico que identifica o órgão ou entidade junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	dcterms:identifier	*foaf:Organization*	*foaf:Agent*	Dublin Core		Alteração de visibilidade
Poder	String	Executivo, legislativo, judiciário ou órgãos essenciais à justiça, como Ministérios Públicos, Defensorias Públicas e Procuradorias		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	dcterms:type	*foaf:Organization*	*foaf:Agent*	Dublin Core		Valores controlados: "Executivo" ou "Legislativo" ou "Judiciário"
Organização Pai	String	Organização à qual o órgão/entidade esteja vinculado		Opcional	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:relation*	foaf:Organization	*foaf:Agent*	Dublin Core		

Imagem / Logo do órgão	Arquivo de imagem	Logomarca oficial do órgão/entidade		Opcional	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:thumbnail* *foaf:logo*	*foaf:Organization*	*foaf:Agent*	Foaf		
Sigla	String	Abreviação que identifica o órgão		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:nick*	*foaf:Organization*	*foaf:Agent*	Foaf		
CNPJ	Numérico	Número designado pela Receita Federal para identificar o órgão/organização		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:identifier*	*foaf:Organization* *vcard:Organization*	*foaf:Agent*	Dublin Core		
UF	String	Unidade federativa em que o órgão/entidade esteja localizado		Obrigatório	Sim	-	*vcard:region*	*foaf:Organization*	*foaf:Agent*	vCard		Valores controlados
Município	String	Município em que o órgão/entidade esteja localizado		Obrigatório	Sim	-	*vcard:locality*	*foaf:Organization*	*foaf:Agent*	vCard		Valores controlados
Logradouro	**String**	**Via pública ou o local onde organização está situada**		**Obrigatório**	**Sim**	**-**	***vcard:street-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Número	**String**	**Numeração**		**Obrigatório**	**Sim**	**-**	***vcard:extend-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Bairro	**String**	**Subárea do município**		**Obrigatório**	**Sim**	**-**	***vcard:extend-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Complemento	**String**	**Informações adicionais do endereço**		**Opcional**	**Sim**	**-**	***vcard:extend-adress***	***foaf:Organization***	***foaf:Agent***	**vCard**		
CEP	**Numérico**	**Código de endereçamento postal**		**Obrigatório**	**Sim**	**-**	***vcard:postalcode***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Site da organização	URL	Sítio eletrônico oficial do órgão/entidade		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:homepage*	*foaf:Organization*	*foaf:Agent*	Foaf		Alteração de visibilidade

Email da organização	Numérico	E-mail institucional do órgão/entidade		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*foaf:mbox* *vcard:hasemail*	*foaf:Organization*	*foaf:Agent*	Foaf ou vCard		Alteração de visibilidade
Telefone de contato da organização	**Numérico**	**Telefone de contato do órgão/entidade**		**Obrigatório**	**Sim**	**Contexto Jurídico-administrativo (Mandate)**	***vcard:tel***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Horário de atendimento ao público	**String**	**Escala de dias e horários de atendimento ao público**		**Obrigatório**	**Sim**	**-**	***vcard:hours***	***foaf:Organization***	***foaf:Agent***	**vCard**		
Status da organização	String	Ativo: o órgão/entidade está ativo junto à Receita Federal do Brasil. Extinto: órgão/entidade não executa mais atividades e estão inativos junto à Receita Federal do Brasil.		Obrigatório	Sim	Contexto Jurídico-administrativo (Mandate)	*dcterms:type*	*foaf:Organization*	*foaf:Agent*	Dublin Core		Valores controlados: "Ativo" ou "Extinto"
Descrição (da organização)	String	Descrição do órgão/entidade, como área de atuação, missão e demais informações que o órgão/entidade achar pertinente compartilhar		Obrigatório	Sim	description.resumo (description.abstract)	*dcterms:description*	*foaf:Organization*	*foaf:Agent*	Dublin Core		
Título do conjunto de dados	String	Nome do conjunto de dados. A nomenclatura deve ser idêntica à que foi elencada no Plano de Dados Abertos – PDA, quando esse instrumento for aplicável e o conjunto tiver sido previsto no	Conjunto de dados	Obrigatório	Sim	Título (Title)	*dcterms:title*	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 2	

		cronograma daquele documento										
Organização publicadora	String	Campo que identifica a organização que está cadastrando o conjunto de dados		Obrigatório	Sim	criador.autor (creator.autor)	*dcterms:publisher*	*dcat:Dataset1*	*dctype:Dataset*	DublinCore	Boas práticas 1, 2 e 5	
Item de inventário de origem	String	Indicar a qual item do inventário o conjunto de dados está relacionado		Opcional	Sim	Fonte (Source)	*dcterms:source*	*dcat:Dataset1*	*dctype:Dataset*	DublinCore	Boas práticas 1, 3 e 5	
Descrição do conjunto de dados	String	Descrição sobre o conjunto de dados, incluindo metadados que o identifique		Obrigatório	Sim	description.resumo (description.abstract)	*dcterms:description*	*dcat:Dataset1*	*dctype:Dataset*	DublinCore	Boas práticas 1 e 2	
Licença de uso	String	Selecionar sob qual espécie de licença o conjunto de dados será publicado		Obrigatório	Sim	Direitos (Rights)	*dcterms:license*	*dcat:Dataset1*	*dctype:Dataset*	DublinCore	Boas práticas 1 e 4	Valores controlados
Área técnica responsável pelo dado	String	Indicar qual área técnica do órgão/entidade é responsável pelo conjunto de dados		Obrigatório	Sim	criador.produtor (creator.productor)	*dcterms:creator*	*dcat:Dataset1*	*dctype:Dataset*	DublinCore	Boas prátocas 1, 2 e 5	Alteração de visibilidade
Email da área técnica responsável	String	Indicar e-mail de contato da área técnica responsável pelo conjunto de dados		Obrigatório	Sim	-	*dcat:contactPoint*	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1, 2 e 29	Alteração de visibilidade
Periodicidade de atualização	String	Indicar qual será a periodicidade com que o conjunto de dados será atualizado		Obrigatório	Sim	data.frequenciaDeAtualização	*dcterms:accrualPeriodicity*	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 21	
Temas aos quais o dado está associado	String	Escolha a área temática à qual o conjunto de dados esteja vinculado		Obrigatório	Sim	Assunto.categoria (subject.category)	*dcat:themeTaxonomy*	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1 e 2	Valores controlados (VCGE)
Assuntos / Palavras-chave	String	Resumo dos principais aspectos		Obrigatório	Sim	assunto.palavra-chave	*dcat:keyword*	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1 e 2	

		do conjunto de dados										
Cobertura temporal início	Numérico	Data ou período de início dos dados disponibilizados		Obrigatório	Sim	abrangência.temporal	dcat:startDate	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1 e 2	Alteração de visibilidade
Cobertura temporal fim	Numérico	Data ou período de fim dos dados disponibilizados		Opcional	Sim	abrangência.temporal	dcat:endDate	*dcat:Dataset1*	*dctype:Dataset*	Dcat	Boas práticas 1 e 21	Alteração de visibilidade
Cobertura espacial	String	Esfera de cobertura espacial do conjunto de dados		Obrigatório	Sim	abrangência.espacial	dcterms:spatial	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 13	Alteração da obrigatoriedade e visibilidade
Granularidade espacial	String	Nível de granularidade espacial do conjunto de dados		Obrigatório	Sim	abrangência.espacial	dcterms:spatial	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 2	Alteração da obrigatoriedade e visibilidade
Descontinuado	String	Indicar se conjunto de dados foi descontinuado e não passará por futuras atualizações		Opcional	Sim	-	*adms:versionNotes*	*dcat:Dataset1*	*dctype:Dataset*	Adms	Boas práticas 1 e 22	Alteração de visibilidade
Data descontinuação	String	Data a partir da qual o conjunto de dados foi descontinuado		Opcional	Sim	data.disponibilidade (date.available) data.validade (date.valid)	*dcterms:valid*	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 22	Alteração de visibilidade
Conjuntos de dados associados (interoperabilidade)	String	Indicar outros conjuntos de dados que estejam associados ao conjunto de dados cadastrado		Opcional	Sim	Relação (Relation)	dcat:qualifiedRelation dcat:dataset *dcterms:relation*	*dcat:Dataset1*	*dctype:Dataset*	Dublin Core ou Dcat	Boas práticas 1 , 3 e 31	Alteração de visibilidade
Colaborador	**String**	**Pessoa ou entidade que contribuiu para a criação, revisão ou manutenção do recurso**		**Opcional**	**Sim**	**Colaborador (Contributor)**	***dcterms:contributor***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 5**	
API habilitada	**String**	**Indica se o recurso possui uma interface de**		**Opcional**	**Sim**	**-**	***dcat:endpointURL***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1, 23 e 35**	

		programação (API) disponível para interação automatizada										
Identificador	**URI / IRI**	**Código ou número único que identifica o recurso ou registro de forma exclusiva**		**Obrigatório**	**Sim**	**Identificador (Identifier)**	***dcterms:identifier***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1 e 9**	
Vocabulário contorlado associado	**String**	**Conjunto padronizado de termos utilizados para descrever ou categorizar os dados**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 15**	
Lista dos Recursos cadastrados	**String**	**Lista os recursos que estão cadastrados no conjunto de dados**		**Opcional**	**Sim**	**-**	***dcat:distribution***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 3**	
Data de publicação	**String**	**Data em que o Conjunto de Dados foi publicado no Portal**		**Automático**	**Sim**	**data.publicação**	***dcterms:issued***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Última alteração	**String**	**Última alteração do conjunto de dados**		**Automático**	**Sim**	**data.alteração**	***dcterms:modified***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Documentação	**String**	**URL de documento que expõe detalhes sobre o conjunto de dados**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 2**	
Granularidade temporal	**String**	**Precisão temporal da cobertura temporal.**		**Opcional**	**Sim**	**abrangência.temporal**	***dcterms:temporal***	***dcat:Dataset1***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1 e 2**	
Metodologia de criação	**String**	**Processo de criação dos dados**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Dataset1***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 2**	
Título de recurso	String	Nome do recurso	Recurso (dados, documentaçã	Obrigatório	Sim	Título (Title)	*dcterms:title*	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1 e 2	

URL do recurso	URL	Endereço da web onde o recurso está hospedado	o, dicionário de dados, API ou outros)	Obrigatório	Sim	Localização (Location)	*dcat:accessURL*	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1, 5 e 20	
Tipo do recurso	String	Selecionar se o recurso se refere a dados, documentação, dicionário de dados, API ou outros		Obrigatório	Sim	Tipo (Type)	*dcterms:type*	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1 e 3	Valores controlados
Formato do recurso	String	Selecionar sob qual formato o recurso será disponibilizado		Obrigatório	Sim	Formato (Format)	*dcterms:format* *dcat:mediaType*	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core ou Dcat	Boas práticas 1 e 3	
Descrição do recurso	String	Breve descrição sobre o recurso		Obrigatório	Sim	description.resumo (description.abstract)	*dcterms:description*	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1 e 2	
Data de criação do recurso	String	Data de criação do recurso		Obrigatório	Sim	**data.criação**	*dcat:startDate*	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1, 2 e 21	
Data de publicação	**String**	**Data em que o recurso foi publicado no Portal**		**Automático**	**Sim**	**data.publicação**	***dcterms:issued***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Última alteração	String	Última alteração do recurso		Automático	Sim	data.alteração	*dcterms:modified*	*dcat:Distribution*	*dcat:Dataset1*	Dublin Core	Boas práticas 1, 2 e 21	
Versão	Numérico	Versão do conjunto de dados que está sendo cadastrada		Opcional	Sim	Identificador (Identifier) identificador.versão	*dcat:hasCurrentVersion*	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1 e 7	Mudança de Conjuntos de Dados para Recurso
Atualização de versão	Numérico	Indicar se houve atualização da versão do conjunto de dados disponibilizado na última alteração realizada		Opcional	Sim	-	*dcat:hasVersion*	*dcat:Distribution*	*dcat:Dataset1*	Dcat	Boas práticas 1 e 8	Mudança de Conjuntos de Dados para Recurso
Notas de versão	**String**	**Observações sobre atualizações ou**		**Opcional**	**Sim**	**-**	***adms:versionNotes***	***dcat:Distribution***	***dcat:Dataset1***	**Adms**	**Boas práticas 1 e 8**	

		alterações na versão										
Versão anterior	**Numério**	**Indica se o recurso tem versão prévia**		**Opcional**	**Sim**	**-**	***dcat:previousVersion***	***dcat:Distribution***	***dcat:Dataset1***	**Dcat**	**Boas práticas 1 e 8**	
Assuntos / Palavras-chave	**String**	**Resumo dos principais aspectos do recurso**		**Obrigatório**	**Sim**	**Assunto (Subject)**	***dcat:keyword***	***dcat:Distribution***	***dcat:Dataset1***	**Dcat**	**Boas práticas 1 e 2**	
Idioma	**String**	**Língua na qual o conteúdo do recurso está escrito ou apresentado**		**Obrigatório**	**Sim**	**Idioma (Language)**	***dcterms:language***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1 e 2**	
Tamanho do arquivo	**Numérico**	**Medida do tamanho do arquivo ou conjunto de dados em bytes**		**Obrigatório**	**Sim**	**format.dimensão**	***dcat:byteSize***	***dcat:Distribution***	***dcat:Dataset1***	**Dcat**	**Boas práticas 1 e 3**	
Sigilo / Direitos de acesso	**String**	**Nível de restrição ou permissões de acesso, indicando quem pode visualizá-lo ou utilizá-lo**		**Opcional**	**Sim**	**Direitos (Rights)**	***dcterms:accessRights***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1 e 2**	
Normativa / Observância legal	**String**	**Normativas / Legislação aplicável**		**Opcional**	**Sim**	**-**	***dcterms:conformsTo***	***dcat:Distribution***	***dcat:Dataset1***	**Dublin Core**	**Boas práticas 1 e 2**	
Título do reúso	String	Nome do reúso por extenso	Reúso	Obrigatório	Sim	Título (Title)	*dcterms:title*	*dcat:Dataset2*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 2	
URL do reúso	URL	Endereço da web onde o reúso está hospedado		Obrigatório	Sim	Localização (Location)	*dcat:accessURL*	*dcat:Dataset2*	*dctype:Dataset*	Dcat	Boas práticas 1 e 5	
Descrição do reúso	String	Explicação sobre o reúso		Obrigatório	Sim	description.resumo (description.abstract)	*dcterms:description*	*dcat:Dataset2*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 2	
Tipo do reúso	String	Selecionar entre as opções fornecidas – aplicativo, site, dados e outros)		Obrigatório	Sim	Tipo (Type)	*dcterms:type*	*dcat:Dataset2*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 3	Valores controlados

Imagem do reúso	Arquivo de imagem	Imagem referente ao reúso		Opcional	Sim	-	*foaf:thumbnail*	*dcat:Dataset2*	*dctype:Dataset*	Foaf	Boas práticas 1 e 2	
Organização a qual o reúso está associado	String	Selecionar qual órgão/entidade a que o reúso está relacionado		Obrigatório	Sim	Relação (Relation) ou Fonte (Source)	*dcterms:relation*	*dcat:Dataset2*	*dctype:Dataset*	Dublin Core	Boas práticas 1 e 5	
Conjuntos de dados de origem	String	Indicar o(s) conjunto(s) de dados que foram utilizados para o reúso		Obrigatório	Sim	Relação (Relation) ou Fonte (Source)	*dcat:Dataset*	*dcat:Dataset2*	*dctype:Dataset*	Dcat	Boas práticas 1, 5, 31 e 35	
Assuntos / Palavras-chave	String	Palavras que destacam o tema do reúso		Obrigatório	Sim	Assunto (Subject)	*dcat:keyword*	*dcat:Dataset2*	*dctype:Dataset*	Dcat	Boas práticas 1 e 2	
Área técnica responsável pelo reúso	**String**	**Setor ou equipe encarregada pela produção ou manutenção do reúso**		**Obrigatório**	**Sim**	**criador.produtor (creator.productor)**	***dcterms:creator***	***dcat:Dataset2***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1 e 5**	**Substituição de "Nome do responsável" por "Área técnica responsável"**
Email da área técnica responsável	**String**	**Endereço de e-mail de contato da área responsável pelo reúso**		**Obrigatório**	**Sim**	**-**	***dcat:contactPoint***	***dcat:Dataset2***	***dctype:Dataset***	**Dcat**	**Boas práticas 1 e 2**	
Data de criação do reúso	**String**	**Data de criação do reúso**		**Obrigatório**	**Sim**	**data.criação**	***dcat:startDate***	***dcat:Dataset2***	***dctype:Dataset***	**Dcat**	**Boas práticas 1, 2 e 21**	
Data de publicação	**String**	**Data em que o recurso foi publicado no Portal**		**Automático**	**Sim**	**data.publicação**	***dcterms:issued***	***dcat:Dataset2***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 21**	
Última alteração	**String**	**Última alteração do reúso**		**Automático**	**Sim**	**data.alteração**	***dcterms:modified***	***dcat:Dataset2***	***dctype:Dataset***	**Dublin Core**	**Boas práticas 1, 2 e 21**	

Fonte: elaborado pela autora (2024).

APÊNDICE B - Sintaxe Turtle da proposta no padrão DCAT

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix adms: <http://www.w3.org/ns/adms#> .

[] a dcat:Catalog ;
        dcterms:title [] ;  # Nome do catálogo de dados
        dcterms:description [] ;  # Descrição do catálogo de dados
        dcat:homepage [] ;  # URL do catálogo de dados
        dcat:dataset []. # Conjuntos de Dados do catálogo

[] a foaf:Organization ;
        foaf:name  []  ;  # Cadastro oficial dos órgãos e entidades da administração direta, autárquica e fundacional do Poder Executivo Federal junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG
        dcterms:description [] ; # Descrição do órgão/entidade, como área de atuação, missão e demais informações que o órgão/entidade achar pertinente compartilhar
        dcterms:type [] ; # Status da organização: Ativo ou Extinto
        dcterms:type [] ; # Tipo de Organização: Governamental ou Não Governamental
        dcterms:type [] ; # Natureza jurídica: Administração direta ou indireta
        dcterms:type [] ; # Esfera: federal, estadual ou municipal
        dcterms:type [] ; # Poder: Executivo, legislativo, judiciário ou órgãos essenciais à justiça, como Ministérios Públicos, Defensorias Públicas e Procuradorias
        dcterms:identifier [] ; # Referencial numérico que identifica o órgão ou entidade junto ao Sistema de Organização e Inovação Institucional do Governo Federal - SIORG
        dcterms:relation [] ; # Organização à qual o órgão/entidade esteja vinculado
        foaf:logo [] ; # Logomarca oficial do órgão/entidade
        foaf:nick [] ; # Sigla / Abreviatura que identifica o órgão
        dcterms:identifier [] ; # CNPJ
```

```
        vcard:Address [
        vcard:region [] ; # Unidade da Federação
        vcard:locality [] ; # Município
        vcard:street-address [] ; # Logradouro
        vcard:extend-address [] ; # Número
        vcard:extend-address [] ; # Bairro
        vcard:extend-address [] ; # Complemento
        vcard:postal-code [] ; # CEP
        ] ;
        foaf:homepage [] ; # Site oficial do órgão/entidade
        foaf:mbox [] ; # Email da organização
        vcard:tel [] ; # Telefone de contato do órgão/entidade
        vcard:hours [] . # Horário de atendimento ao público

[] a dcat:Dataset ;
        dcterms:title [] ; # Nome do conjunto de dados
        dcterms:description [] ; # Descrição do conjunto de dados, inclusive metadados que o identifiquem
        dcat:keyword [] ; # Principais aspectos do conjunto de dados
        dcterms:publisher [] ; # Organização que está cadastrando o conjunto de dados
        dcterms:source [] ; # Inventário de origem do conjunto de dados
        dcterms:license [] ; # Licença o conjunto de dados será publicado
        dcterms:creator [] ; # Área técnica é responsável pelo conjunto de dados
        dcat:contactPoint [] ; # Email de contato da área técnica responsável pelo conjunto de dados
        dcterms:accrualPeriodicity [] ; # Periodicidade de atualização do conjunto de dados
        dcat:themeTaxonomy [] ; # Área temática à qual o conjunto de dados está vinculado
        dcat:startDate [] ; # Data ou período de início dos dados disponibilizados
        dcat:endDate [] ; # Date ou período de fim dos dados disponibilizados
        dcterms:spatial [] ; # Esfera de cobertura espacial do conjunto de dados
        dcterms:spatial [] ; # Nível de granularidade espacial do conjunto de dados
```

```
        adms:notes [] ; # Indicar se conjunto de dados foi descontinuado
        dcterms:valid [] ; # Data de descontinuação do conjunto de dados
        dcat:qualifiedRelation [] ; # Indicar outros conjuntos de dados associados
        dcterms:contributor [] ; # Pessoa ou entidade que contribuiu para a criação, revisão ou manutenção do recurso
        dcat:endpointURL [] ; # Indica se o recurso possui uma interface de programação (API) disponível para interação automatizada
        dcterms:identifier [] ; # Código ou número único que identifica o recurso ou registro de forma exclusiva
        dcterms:conformsTo [] ; # Vocabulário controlado associado
        dcat:distribution [] . # Lista os recursos cadastrados no conjunto de dados

[] a dcat:Distribution ;
        dcterms:title [] ; # Nome do recurso
        dcat:accessURL [] ; # Endereço da web onde o recurso está hospedado
        dcterms:type [] ; # Tipo do recurso (Dados, documentação, dicionário de dados, API)
        dcterms:format [] ; # Formato de disponibilização do recurso
        dcterms:description [] ; # Descrição do recurso
        dcat:startDate [] ; # Data de criação do recurso
        dcat:issued [] ; # Data de publicação do recurso
        dcterms:modified [] ; # Data de alteração do recurso
        dcat:hasCurrentVersion [] ; # Versão do conjunto de dados
        dcat:hasVersion [] ; # Indicar se houve atualização de versão na última modificação
        adms:versionNotes [] ; # Observações sobre atualizações ou alterações na versão
        dcat:previousVersion [] ; # Indica versões prévias do recurso
        dcat:keyword [] ; # Principais aspectos do recurso
        dcterms:language [] ; # Língua na qual o conteúdo do recurso está escrito ou apresentado
        dcat:byteSize [] ; # Tamanho do recurso em bytes
        dcterms:accessRights [] ; # Nível de restrição ou permissões de acesso, indicando quem pode visualizá-lo ou utilizá-lo
        dcterms:conformsTo [] . # Normativas / Legislação aplicável
```

```
[] a dcat:Dataset ;
        dcterms:title [] ; # Nome do reúso
        dcat:accessURL [] ; # Endereço da web onde o reúso está hospedado
        dcterms:description [] ; # Descrição do reúso
        foaf:thumbnail [] ; # Imagem referente ao reúso
        dcterms:relation [] ; # Organização a qual o reúso está associado
        dcat:dataset [] ; # Indicar qual(is) conjunto(s) de dados utilizados para o reúso
        dcat:keyword [] ; # Termos que identificam o reúso
        dcat:startDate [] ; # Data de criação do reúso
        dcat:issued [] ; # Data de publicação do reúso
        dcterms:modified [] ; # Data de alteração do reúso
        dcterms:creator [] ; # Setor ou equipe responsável pela produção ou manutenção do reúso
        dcat:contactPoint [] . # Endereço de email da área responsável pelo reúso
```

ANEXO A - Elementos do e-PMG 1.1

Apresenta-se no Quadro 23 os 20 metadados do e-PMG 1.1, seus respectivos identificadores, qualificadores, definições, definições de obrigatoriedade e objetivos:

Quadro 23 - Elementos do e-PMG 1.1

ID	Nome	Identificador	Qualificadores		Definição	Obrigatoriedade	Objetivo
1	Abrangência	*Coverage*	Abrangência.espacial (*coverage.spatial)*	Valores estruturados: • Código postal	Extensão espacial e temporal do recurso	Obrigatório se aplicável	Permitir que uma pesquisa se restrinja a recursos sobre um determinado lugar ou tempo
			Abrangência.temporal (coverage.temporal)	Valores estruturados: • Datá de ínicio • Data Fim			
2	Assunto	*Subject*	assunto.categoria (*subject.category*)		Palavras-chaves ou termos que representem corretamente o conteúdo do recurso	Categoria – Obrigatório Palavras-chave – Opcional Pessoa – Opcional	Possibilitar a recuperação do recurso com base em termos de interesse à pesquisa. Indicar de forma sucinta o conteúdo geral do recurso
			assunto.palavra-chave (*subject.keyword*)				
			Assunto.pessoa (*subject.person*)				
3	Colaborador	*Contributor*	Colaborador.papel (*contributor.role)*		Entidade (pessoa ou organização) que contribuiu para a elaboração do conteúdo do recurso ou para sua divulgação	Opcional	Permitir que usuários pesquisem e localizem uma pessoa ou organização que contribuiu para o recurso
4	Contexto jurídico-administrativo	*Mandate*	contexto.estatutoDeAutorização (*mandate.authorisingStatute*)		Legislação e normas que regulam e autorizam o funcionamento e o desempenho das atividades da organização responsável pela produção e/ou disponibilização do recurso	Opcional	Fornecer informações sobre origem, desenvolvimento e atividades da organização responsável pelo recurso a fim de torná-lo mais compreensível

5	Criador	*Creator*	criador.autor *(creator.autor)*	Entidade (pessoa ou organização) responsável pelo conteúdo intelectual do recurso	Uma entidade (pessoa ou organização) originalmente responsável pela criação e manutenção do conteúdo do recurso	Obrigatório	Permitir ao usuário encontrar recursos que foram criados por uma pessoa ou organização. Demonstrar a autenticidade de um recurso, indicando a entidade diretamente responsável pela sua criação
			criador.produtor (*creator.productor*)	Nome da entidade (pessoa ou organização) responsável pela produção e manutenção dos documentos arquivísticos			
6	Data	*Date*	data.aceitação (*date.accepted*)	Data de aceitação de um recurso, como uma tese aceita por uma universidade ou um artigo aceito para uma publicação	Uma data associada a um evento no ciclo de vida do recurso	Obrigatório	Indicar a data de um recurso. Permitir que um usuário localize um recurso, especificando uma data
			data.alteração (*date.modified*)	Data na qual o recurso foi alterado			
			data.criação (*date.created*)	Data da criação ou produção do recurso			
			data.capturado (*date.declared*)	Data em que o recurso é integrado por um sistema informatizado			
			data.disponibilidade (*date.available*)	Data em que o recurso esteve ou estará disponível para acesso			
			data.publicação (*date.issued*)	Data da publicação do recurso			
			data.freqüênciaDeAtualização (*date.updatingFrequency*)	Indica a frequência em que o recurso é atualizado			
			data.próximaVersão (*date.nextVersionDue*)	Data em que houve a substituição do recurso			
			data.recebimento (*date.acquired*)	Data em que o recurso deu entrada na organização			
			data.submissão (*date.submitted*)	Data em que um recurso é submetido à apreciação ou julgamento			

			Data.validade (*date.valid*)	Data ou período em que um recurso é válido			
7	Descrição	*Description*	Descrição.resumo (*description.abstract*)	Exposição sucinta do conteúdo do recurso	Resumo e explicação do conteúdo do recurso	Opcional	Permitir que o usuário avalie a relevância do recurso
			Descrição.sumário (*description.tableOfContents*)	Lista de subunidades do conteúdo do recurso			
8	Destinação	*Disposal*	destinação.dataDeAuto-Remoção (*disposal.autoRemoveDate*)	Proibido para documentos arquivísticos. A data de auto-remoção é usada para remoções automáticas onde não há necessidade de intervenção humana ou avaliação	Decisão quanto ao encaminhamento dos documentos arquivísticos para guarda permanente ou eliminação	Obrigatório para documentos arquivísticos; Opcional para outros tipos de recurso	Apoiar o gerenciamento do ciclo de vida dos recursos e alertar os gestores de documentos e arquivistas para as datas de avaliação dos recursos
			destinação.ação (*diposal.action*)	A ação a ser efetivada quando uma determinada condição for alcançada. As ações podem ser: eliminação, quando o recurso cumprir com o prazo previsto para sua guarda; exportação, quando for previsto a sua transferência ou recolhimento para uma instituição de preservação.			
			destinação.prazoDeGuarda (*disposal.time.Period*)	Prazo em que o recurso deve ser mantido. Em alguns casos a contagem desse prazo inicia-se após um evento especificado no metadado Condições de destinação			
			destinação.condições (*disposal.conditions*)	Um evento que dispara a contagem do prazo de guarda. Por exemplo, o encerramento de um dossiê			
			destinação.data (*disposal.date*)	A data que uma ação de destinação deve ser efetivada, isto é, a data em que o recurso cumpriu o prazo de guarda contado a partir do evento Condições de destinação			

			destinação.responsável (*disposal.authorizedBy*)	A identidade da função, pessoa ou política responsável pela autorização da destinação			
			destinação.nota (*disposal.comment*)	Comentários sobre a destinação			
			destinação.localDeDestinação (*disposal.exportDestination*)	O local previsto para destinação de um recurso exportado, como por exemplo, documentos que serão recolhidos ao Arquivo Nacional para guarda permanente			
			destinação.estágioDeExportação (*disposal.exportStatus*)	Informação sobre o andamento da exportação, como por exemplo, se está pendente, se já foi executado ou se está falho			
9	Destinatário	*Addressee*	Destinatário.cópia (*addressee.copy*)	O(s) nome(s) da(s) entidades para quem foi encaminhado uma cópia do recurso	Entidade (pessoa ou organização) para quem a informação contida no documento foi dirigida	Obrigatório se aplicável	Identificar a entidade (organização ou pessoa) a quem o recurso foi dirigido
10	Direitos	*Rights*	direitos.direitosAutorais (Obrigatório de aplicável)	Indicação de propriedade legal e dos direitos a de uso e reutilização do todo ou de parte do recurso	Informação relativa às restrições e permissões sobre o acesso e o uso ao recurso, tais como ler, copiar, distribuir, publicar ou ainda fazer uso de todo ou de parte do recurso	Opcional	Indicar ao usuário as restrições quanto ao acesso e uso do recurso e informar quando essas restrições podem ser alteradas ou removidas. Facilitar a gestão apropriada de recursos sensíveis ou sigilosos
			direitos.classificaçãoDoGrauDeSigilo (Obrigatório)	Informar se o recurso é ostensivo ou sigiloso. No caso de recurso sigiloso, deve-se identificar o nível de sigilo. Valores permitidos para o qualificador: - Sem classificação (acesso ostensivo) - reservado - confidencial - secreto - ultra-secreto			
			Direitos.descritor (*rights.descriptor*) (Opcional)	Detalhamento do significado de uma marcação de proteção, como por exemplo, Contratos,			

				Pessoal, Segurança do estado.			
			Direitos.dataDeExpiração DoGrauDeSigilo (Obrigatório se aplicável)	Data que expira o grau de sigilo, a partir da qual pode ser reclassificado ou desclassificado			
			Direitos.custodiador (*rights.custodian*) (Opcional)	Identificação do usuário ou da função com poderes de gestão sobre o recurso			
			Direitos.listaDeAcessoDo sGrupos (rights.groupAccessList) (Opcional)	Nome de um grupo ou grupos que têm o direito de ter acesso ao recurso			
11	Fonte	*Source*	-		Identificação do recurso que deu origem ao recurso atual	Opcional	Encontrar recursos que foram desenvolvidos a partir do conteúdo de uma fonte específica
12	Formato	*Format*	formato.formatoDeArquiv o (*format.fileFormat*)	Formato corrente do arquivo em que o recurso está registrado	Conjunto de características físicas e lógicas que define a apresentação do recurso	Opcional	Identificar formatos específicos do recurso. Informar sobre a dependência de hardware e software necessários para apresentar o recurso.
			Format.dimensão (*format.extent*)	tamanho físico e/ou capacidade do recurso			
			Formato.mídia (*format.medium*)	Material ou carregador físico do recurso como CD-ROM; Papel; Disco Rígido			
13	Identificador	*Identifier*	identificador.idDoSistema (*identifier.systemID*)	Número gerado automaticamente e atribuído pelo sistema informatizado, quando o recurso é criado pela primeira vez	Uma referência inequívoca ao recurso em um determinado contexto	Identificador do recurso: Obrigatório Código de classificação: Obrigatório para documentos arquivísticos Versão: Obrigatório se aplicável	Identificar de forma unívoca e persistente um recurso
			identificador.códigoDeCla ssificação (*identifier.fileplanID*)	Classificação atribuída ao documento arquivístico com base em um plano de classificação			
			identificador.versão	Identificação da versão ou edição do recurso para indicar a existência			

				de versões anteriores ou posteriores ao recurso descrito			
14	Idioma	*Language*	-		Idioma em que é expresso o conteúdo do documento	Opcional	Identificar o idioma do conteúdo intelectual do recurso. Permitir a pesquisa a um determinado idioma
15	Localização	*Location*	-		Local de armazenamento atual do recurso	Opcional	Facilitar o acesso ao recurso. Gerenciar o armazenamento.
16	Preservação	*Preservation*	preservação.formatoOriginal (*preservation.originaFormat*)	Nome do formato em que o recurso foi primeiramente criado	Informação para apoiar o acesso, uso e preservação de longo prazo de um recurso	Opcional	Apoiar as atividades de preservação do recurso a fim de permitir o seu acesso e uso
			preservação.dependênciaDeHardware	Hardware necessário para acesso ao recurso			
			preservação.dependênciaDeSoftware	Software e versão necessários para acesso ao recurso			
			preservação.DependênciaDeSistemaOperacional	Sistema operacional e versão necessários para acesso ao recurso			
			preservação.resolução	Resolução de imagem digital fixa ou em movimento em pixels			
			preservação.dataDaAção	Data e hora em que a ação de preservação foi realizada			
			preservação.tipoDeAção	A ação de preservação realizada. Valores: conversão, atualização, microfilmagem.			
			preservação.descriçãoDaAção	Detalhes específicos da ação de preservação, incluindo o estado original, as mudanças realizadas, os motivos das mudanças e a autorização para as mudanças.			

			preservação.ferramentasDaAção	Ferramentas utilizadas na ação de preservação			
			preservação.proximaAção	Próxima ação ou avaliação de preservação pela qual a recurso deve passar			
			preservação.dataDaProximaAção	A data programada para a execução da próxima ação. Deve ser utilizado em conjunto com historiaProximaAção			
17	Publicador	*Publisher*	-		Obrigatório se aplicável	Uma organização ou pessoa responsável por tornar o recurso disponível, o qual pode ser uma edição impressa ou uma publicação eletrônica	Identificar o responsável pela disponibilização do recurso
18	Relação	*Relation*	relação.conformidadeCom (*relation.conformsTo*)	Uma referência a um padrão estabelecido ao qual o recurso está em conformidade	Uma referência a um outro recurso relacionado	Opcional	Registrar o tipo de relação entre os recursos. Assegurar a identificação de todos os recursos associados a fim contextualizar o recurso que está sendo descrito
			relação.descriçãoDaRelação	Explicação da relação entre dois ou mais recursos ou fontes de informação			
			relação.éFormatoDe (relation.isFormatOf)	Referência ao recurso que deu origem ao recurso descrito			
			relação.éParteDe (relation.isPartOf)	O recurso descrito é parte física ou lógica do recurso referenciado. Aplicável apenas a documentos ou partes de documentos e nunca a citações ou referências			
			relação.éReferenciadoPor (relation.isReferencedBy)	O recurso descrito é referenciado, citado, ou apontado pelo recurso relacionado, como uma lei que é citada em um ofício			
			relação.éRequeridoPor (relation.isRequiredBy)	O recurso descrito é requerido para apoiar sua			

				utilização ou manter a coerência de conteúdo de um outro recurso, ou ainda para interpretar o recurso, como um programa que é requerido para visualizar um documento em PDF			
			relação.éSubstituídoPor (relation.isReplacedBy	O recurso descrito foi substituído, revogado ou tornado obsoleto por outro recurso como uma lei que é revogada e substituída por outra			
			relação.éVersãoDe (relation.isVersionOf)	O recurso descrito é uma versão, edição ou adaptação do recurso referenciado. Uma mudança na versão implica em mudanças no conteúdo. Inclui traduções do recurso. Deve-se relacionar o recurso descrito com outras versões, edições ou adaptações que lhe sejam anteriores.			
			relação.razãoDaTarjaOuTruncamento	A razão para a publicação ou emissão de um truncamento ou extrato, como privacidade ou sigilo			
			relação.referência (relation.references)	O recurso descrito referencia, cita ou de alguma forma aponta para o recurso referenciado. Deve-se indicar as citações ou apontadores que o recurso descrito estabelece com outro recurso, como um ofício cita uma lei			
18	Relação	*Relation*	relação.requer (relation.requires)	O recurso necessita de outro recurso para assegurar a sua inteligibilidade, ação ou comunicação, como um documento em PDF requer um programa para visualização.			

			relação.substitui (relation.replaces)	O recurso substitui, revoga ou torna obsoleto um outro recurso como uma lei que substituiu uma anterior.			
			relação.tarjaOuTruncamento	O recurso descrito tem uma versão com alguma parte do conteúdo tarjado ou removido			
			relação.temFormatoDe (relation.hasFormat)	Referência à reprodução do recurso em outro formato.			
			relação.temParteDe (relation.haspart)	Recursos individuais que estão incluídos física ou logicamente no recurso descrito. Este qualificador é aplicável apenas a documentos ou partes de documentos e nunca a citações ou referências aos mesmos.			
			relação.temVersãoDe (relation.hasVersion)	O recurso descrito tem versões, edições, traduções ou adaptações posteriores, como versão 4.0 de um documento ou 5ª. edição de um livro			
19	Tipo	Type	-		Natureza ou gênero do conteúdo do recurso	Obrigatório se aplicável	Informar sobre a tipologia do recurso. Apoiar a interpretação do conteúdo do recurso
20	Título	Title	título.tituloAlternativo (title.alternativeTitle)	É um substitutivo ou uma outra forma de designar um recurso, podendo ser em outro idioma	Palavra, frase ou grupo de caracteres que nomeia um recurso e que pode corresponder à sua denominação oficial ou legal, ou ainda ao nome que aparece de forma evidente ou explícita	Título - Obrigatório Título alternativo - Opcional	Permitir a localização de um recurso por meio de um título. Servir como elemento de acesso para a pesquisa. Possibilitar também a realização de pesquisas mais precisas.

Fonte: Adaptado de Brasil (2014)